Vol. XXXIV. Parts III and IV December, 1947

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

HARALD CRAMÉR — J. B. S. HALDANE
R. C. GEARY — G. M. MORANT
MAJOR GREENWOOD — JOHN WISHART

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

Reprinted by offset-litho 1964

[*Issued* 30 *December*, 1947]

VOLUME XXXIV, PARTS III AND IV DECEMBER 1947

ON THE DISTRIBUTION OF THE RANK CORRELATION COEFFICIENT τ WHEN THE VARIATES ARE NOT INDEPENDENT

BY WASSILY HÖFFDING

I. INTRODUCTION

1. Consider a population distributed according to two variates x, y. Two members (x_1, y_1) and (x_2, y_2) of the population will be called *concordant* if both values of one member are greater than the corresponding values of the other one, that is if

$$x_1 < x_2,\ y_1 < y_2 \quad \text{or} \quad x_1 > x_2,\ y_1 > y_2.$$

They will be called *discordant* if for one member one value is greater and the other one smaller than for the other member, that is if

$$x_1 < x_2,\ y_1 > y_2 \quad \text{or} \quad x_1 > x_2,\ y_1 < y_2.$$

The probability p that two members drawn from the population at random without replacement are concordant will be called the *probability of concordance*, the probability q that they are discordant will be called the *probability of discordance*.

In the following only populations will be considered for which the probabilities of $x_1 = x_2$ or $y_1 = y_2$ are zero, so that

$$p + q = 1. \tag{1}$$

The main types of such populations are (a) an infinite population with both x and y distributed continuously, (b) a finite population where all values of x and all values of y are different among themselves. The condition that the two members are drawn without replacement is, of course, only relevant in case (b).

For a sample of n members drawn from the population, the probabilities of concordance and discordance are defined in the same manner as for the population. They will be denoted by p' and q' to distinguish them from the population values. If for the population (1) is fulfilled, it may be assumed that all values of x and all values of y in the sample are different, so that

$$p' + q' = 1. \tag{2}$$

It follows from the definition that p' is the relative frequency of concordant pairs among the $\binom{n}{2}$ pairs which can be formed from the members of the sample.

The probability of concordance expresses an essential property of a bivariate distribution. It may in itself be considered as a measure of correlation. p' is an estimate of p; it will be shown that the mean value of p' is p. If a coefficient lying between the limits -1 and $+1$ is preferred, the quantity

$$\tau = p' - q' = 2p' - 1 \tag{3}$$

may be taken.

2. The quantity p here termed the probability of concordance was apparently first considered by Esscher (1924) who also used the quantity

$$D = \frac{1}{\binom{n}{2}} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \operatorname{sign}(x_i - x_j) \operatorname{sign}(y_i - y_j),$$

(where x_i, y_i, $i = 1, \ldots, n$, are the sample values of the variates) which is the same as the coefficient τ as defined by (3). Esscher showed that if x and y are normally correlated with correlation coefficient r, the expectation of $D = \tau$ is

$$E(\tau) = \frac{2}{\pi}\sin^{-1} r. \tag{4}$$

Hence, from this equation, he suggested estimating r from ranked data by means of the relation

$$r = \sin\frac{\pi}{2}\tau = \sin\pi(p' - \tfrac{1}{2}).$$

For the variance of τ Esscher found in the case of a normally distributed population

$$\frac{1}{4}\binom{n}{2}\operatorname{var}(\tau) = pq + \frac{n-2}{2}\left\{\frac{1}{9} - \left(\frac{2}{\pi}\sin^{-1}\left(\frac{r}{2}\right)\right)^2\right\}, \tag{5}$$

where

$$4pq = 1 - \left(\frac{2}{\pi}\sin^{-1} r\right)^2.$$

While Esscher saw in p' and $D = \tau$ only a means for estimating r, Lindeberg (1926) stressed the significance of the probability of concordance itself for judging the degree of dependence between the variates. He proposed for that purpose the coefficient

$$P = 100p' - 50 = 50\tau,$$

called by him *Korrelationsprozent*. Lindeberg also gave, without proof, a formula for the variance of p' in the general case of correlated variates (see (13) below).

Jordan (1927) suggested using, instead of Lindeberg's P, the coefficient later termed by Kendall τ.

Kendall (1938), independently of the above authors, proposed τ as a measure of rank correlation. He completely solved the problem of the sampling distribution of τ in a universe in which all possible rankings are equally probable, showing that it rapidly tends to normality for increasing n.

3. The main object of this paper is to show that the sampling distribution of p' (and hence that of τ) tends to normality as $n \to \infty$ for any population with continuously distributed x and y if a certain condition is fulfilled (Part IV). In addition, Lindeberg's formula for the variance of p' is proved (Part II) and extended for a finite population (Part V). Finally, in Part VI the problem of estimating $\operatorname{var}(p')$ from the sample is considered.

II. Mean value and variance of p' in the case of an infinite population

4. Consider a sample of n drawn at random from an infinite population with continuous x and y. Replace the values of x and of y in the sample by their ranks and arrange the members of the sample so that the ranking of x is $1, 2, \ldots, n$. Then the ranking of y is a permutation

$$\Pi = (\pi_1, \ldots, \pi_n)$$

of the numbers $1, \ldots, n$.

Let I and J be the numbers of inversions in the permutations $(\pi_n, \pi_{n-1}, \ldots, \pi_1)$ and $(\pi_1, \ldots, \pi_n)$. Then

$$p' = \frac{2I}{n(n-1)}, \quad q' = \frac{2J}{n(n-1)}. \tag{6}$$

Thus the knowledge of the permutation Π corresponding to the given sample is sufficient for evaluating p'.

5. Let $P(\Pi)$ be the probability of drawing a random sample represented by the permutation Π. Let $p'(\Pi)$ be the probability of concordance for such a sample. Then

$$p = \Sigma P(\Pi)\, p'(\Pi), \tag{7}$$

where the sum is extended over all permutations Π of n numbers.

The right-hand side of (7) is equal to the mean value of p'. Hence

$$Ep' = p. \tag{8}$$

Consider, in generalization of p', the probability w' that among $m \leqslant n$ members drawn from the sample at random without replacement, certain pairs of members are concordant; for instance, among four members A, B, C, D, the pairs AB, AC, AD; or the pairs AB, CD, etc. Let w be the corresponding probability for the parent population. Then it is seen in the same manner as with p' that

$$Ew' = w. \tag{9}$$

Thus, if we can express $(p')^\mu$, the probability of drawing μ concordant pairs from the sample, replacing each pair after drawing it, by probabilities without replacement of the type w', we can also, in virtue of (9), represent $E(p')^\mu$ by population parameters of the type w.

6. Now, $(p')^2$, the probability of drawing from the sample one concordant pair and, after replacing it, of drawing again a concordant pair, is the sum of the following three probabilities:

(a) the probability of getting the same pair in both drawings $\left(1 \Big/ \binom{n}{2}\right)$, multiplied by the probability that this pair is concordant (p');

(b) the probability that the second pair has one member in common with the first pair $\left(2(n-2) \Big/ \binom{n}{2}\right)$, multiplied by the probability, say k', that among three members A, B, C drawn from the sample without replacement, one, say A, is concordant with the other two;

(c) the probability that the second pair has no member in common with the first one $\left(\binom{n-2}{2} \Big/ \binom{n}{2}\right)$, multiplied by the probability that among four members A, B, C, D drawn without replacement, two pairs without a member in common, say AB and CD, are concordant. The latter probability may be denoted by $(p^2)'$ since the corresponding probability for the infinite population is p^2.

Thus,
$$\binom{n}{2}(p')^2 = p' + 2(n-2)\,k' + \binom{n-2}{2}(p^2)', \tag{10}$$

and, applying (9),
$$\binom{n}{2} E(p')^2 = p + 2(n-2)\,k + \binom{n-2}{2} p^2. \tag{11}$$

Hence, we have for the variance of p'

$$\binom{n}{2} \operatorname{var}(p') = \binom{n}{2}\{E(p')^2 - p^2\} = p + 2(n-2)\,k - (2n-3)\,p^2 \tag{12}$$

or
$$\binom{n}{2} \operatorname{var}(p') = \binom{n}{2} \mu_2(p') = p(1-p) + 2(n-2)\,(k-p^2). \tag{13}$$

This is identical with the formula given without proof by Lindeberg (1926).

7. In the case considered by Kendall where all permutations Π of n numbers are equally probable, the permutations of $m \leqslant n$ also are equiprobable. Hence

$$p = P(1,2) = q = P(2,1) = \tfrac{1}{2}.$$

Further, representing k as the mean value of k' in a sample of 3, we find

$$k = P(123) + \tfrac{1}{3}P(132) + \tfrac{1}{3}P(213) = (1 + \tfrac{1}{3} + \tfrac{1}{3})\frac{1}{3!} = \frac{5}{18}.$$

Inserting these values in (13), we have

$$\operatorname{var}(p') = \frac{2n+5}{18n(n-1)},$$

$$\operatorname{var}(\tau) = 4\operatorname{var}(p') = \frac{2(2n+5)}{9n(n-1)}$$

in accordance with Kendall's formula.

III. Some algebraic formulae

8. We shall now consider some algebraic relations to be used in the proof of normality of p' for large n.

Let $f_d(\rho)$ be a polynomial of degree d in ρ. Then

$$\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho} f_d(\rho) = \begin{cases} 0 & \text{if } d < \beta, \\ a_0\beta! & \text{if } d = \beta, \end{cases} \tag{14}$$

where a_0 is the coefficient of the highest power ρ^d in $f_d(\rho)$.

To prove (14) write $\qquad f_d(\rho) = a_0\rho^{[d]} + a_1\rho^{[d-1]} + \dots,$

where $\qquad \rho^{[0]} = 1, \quad \rho^{[\delta]} = \rho(\rho-1)\dots(\rho-\delta+1), \quad (\delta \geqslant 1).$

Then (14) follows from the fact that

$$\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}\rho^{[\delta]} = \beta^{[\delta]}\sum_{\rho-\delta=0}^{\beta-\delta} (-1)^{\beta-\rho}\binom{\beta-\delta}{\rho-\delta}$$

is equal to $\beta^{[\delta]}(1-1)^{\beta-\delta} = 0$ if $\beta - \delta > 0$ and to $\beta!$ if $\beta = \delta$.

9. For any non-negative integer ν we may write

$$n^{\nu} = d_{\nu 0}^{(\alpha)}(n-\alpha)^{[\nu]} + d_{\nu 1}^{(\alpha)}(n-\alpha)^{[\nu-1]} + \dots + d_{\nu,\nu-1}^{(\alpha)}(n-\alpha) + d_{\nu\nu}^{(\alpha)}. \tag{15}$$

We will study certain properties of the coefficients $d_{\nu\kappa}^{(\alpha)}$.

From (15) it is seen immediately that

$$d_{\nu 0}^{(\alpha)} = 1. \tag{16}$$

Inserting in (15) $n = \alpha + \beta$ $(\beta = 0, 1, \dots)$ we have

$$(\alpha+\beta)^{\nu} = d_{\nu,\nu-\beta}^{(\alpha)}\beta! + d_{\nu,\nu-\beta+1}^{(\alpha)}\beta^{[\beta-1]} + \dots + d_{\nu,\nu-1}^{(\alpha)}\beta + d_{\nu\nu}^{(\alpha)} \tag{17}$$

or

$$d_{\nu\nu}^{(\alpha)} = \alpha^{\nu}, \quad d_{\nu,\nu-\beta}^{(\alpha)} = \frac{1}{\beta!}\left\{(\alpha+\beta)^{\nu} - \sum_{\lambda=0}^{\beta-1}\beta^{[\lambda]}d_{\nu,\nu-\lambda}^{(\alpha)}\right\}, \quad (\beta = 1, 2, \dots). \tag{18}$$

Hence we find by induction

$$d_{\nu,\nu-\beta}^{(\alpha)} = \frac{1}{\beta!}\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}(\alpha+\rho)^{\nu}. \tag{19}$$

If we take this as definition of $d_{\nu\kappa}^{(\alpha)}$ for $\kappa < 0$, we have in virtue of (14)

$$d_{\nu\kappa}^{(\alpha)} = 0 \quad \text{for} \quad \kappa < 0. \tag{20}$$

Expanding $(\alpha+\rho)^{\nu}$ we have from (19)

$$d_{\nu,\nu-\beta}^{(\alpha)} = \frac{1}{\beta!}\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}\sum_{\sigma=0}^{\nu}\binom{\nu}{\sigma}\alpha^{\sigma}\rho^{\nu-\sigma} = \sum_{\sigma=0}^{\nu}\binom{\nu}{\sigma}\alpha^{\sigma}\frac{1}{\beta!}\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}\rho^{\nu-\sigma}.$$

Comparing the last sum with (19) and writing

$$d_{\nu\kappa} = d^{(0)}_{\nu\kappa},$$

we have

$$d^{(\alpha)}_{\nu,\nu-\beta} = \sum_{\sigma=0}^{\nu} \binom{\nu}{\sigma} d_{\nu-\sigma,\nu-\beta-\sigma}\alpha^{\sigma}$$

or, putting $\nu-\beta=\kappa$ and noting that, by (20), $d_{\nu-\sigma,\kappa-\sigma}=0$ for $\sigma>\kappa$,

$$d^{(\alpha)}_{\nu\kappa} = \sum_{\sigma=0}^{\kappa} \binom{\nu}{\sigma} d_{\nu-\sigma,\kappa-\sigma}\alpha^{\sigma}. \tag{21}$$

We have the recurrence relation

$$d^{(\alpha)}_{\nu+1,\kappa} - d^{(\alpha)}_{\nu\kappa} = (\alpha+\nu+1-\kappa)\, d^{(\alpha)}_{\nu,\kappa-1} \tag{22}$$

which can be obtained by multiplying (15) by n, then writing down (15) with $\nu+1$ instead of ν, and comparing coefficients in both expressions.

10. We prove now two properties of the coefficients $d^{(\alpha)}_{\nu\kappa}$.

(I) $d_{\nu\kappa}$ is a polynomial in ν of degree 2κ, the term of highest degree being $\nu^{2\kappa}/2^{\kappa}\kappa!$.

In virtue of (16) this is true for $\kappa=0$. And if it is true for $\kappa-1$, the highest term of $d_{\nu+1,\kappa}-d_{\nu\kappa}$ is, by (22) with $\alpha=0$, $\nu^{2\kappa-1}/2^{\kappa-1}(\kappa-1)!$, and hence that of $d_{\nu\kappa}$, by a well-known theorem,

$$\frac{1}{2\kappa}\frac{1}{2^{\kappa-1}(\kappa-1)!}\nu^{2\kappa} = \frac{1}{2^{\kappa}\kappa!}\nu^{2\kappa}.$$

(II) $d^{(\gamma-t}_{t-\rho,\kappa}$ is a polynomial in t of degree 2κ with the highest term $(-1)^{\kappa}t^{2\kappa}/2^{\kappa}\kappa!$.

From (21),

$$d^{(\gamma-t)}_{t-\rho,\kappa} = \sum_{\sigma=0}^{\kappa} \binom{t-\rho}{\sigma} d_{t-\rho-\sigma,\kappa-\sigma}(\gamma-t)^{\sigma}.$$

In $\binom{t-\rho}{\sigma}$ the highest term in t is $\frac{1}{\sigma!}t^{\sigma}$.

In $d_{t-\rho-\sigma,\kappa-\sigma}$ the highest term in t is $\frac{1}{2^{\kappa-\sigma}(\kappa-\sigma)!}t^{2\kappa-2\sigma}$ (by (I)).

In $(\gamma-t)^{\sigma}$ the highest term in t is $(-1)^{\sigma}t^{\sigma}$.

Hence, in $d^{(\gamma-t}_{t-\rho,\kappa}$ the highest term is

$$\sum_{\sigma=0}^{\kappa} \frac{(-2)^{\sigma}}{2^{\kappa}\sigma!(\kappa-\sigma)!}t^{2\kappa} = \frac{1}{2^{\kappa}\kappa!}(1-2)^{\kappa}t^{2\kappa} = \frac{(-1)^{\kappa}}{2^{\kappa}\kappa!}t^{2\kappa}.$$

11. $d_{\nu,\nu-\beta}$ has also a combinatorial meaning.

Let

$$\Sigma_{\beta}(\nu) = \Sigma\frac{\nu!}{\nu_1!\,\nu_2!\ldots\nu_{\beta}!}, \quad \Sigma'_{\beta}(\nu) = \Sigma'\frac{\nu!}{\nu_1!\,\nu_2!\ldots\nu_{\beta}!}$$

where Σ indicates summation over all $\nu_i \geqslant 0$, Σ' over all $\nu_i \geqslant 1$, and in both cases $\nu_1+\ldots+\nu_{\beta}=\nu$. $\Sigma_{\beta}(\nu)$ is the number of ways of allocating ν objects on β places, and $\Sigma'_{\beta}(\nu)$ is the number of ways of allocating ν objects on β places in such a manner that no place remains empty.

We have

$$\Sigma_{\beta}(\nu) = \beta^{\nu},$$

and a little consideration shows that

$$\Sigma_{\beta}(\nu) = \Sigma'_{\beta}(\nu) + \binom{\beta}{1}\Sigma'_{\beta-1}(\nu) + \ldots + \binom{\beta}{\beta-1}\Sigma'_{1}(\nu).$$

Comparing this with (17) we see that

$$d_{\nu,\nu-\beta} = \frac{1}{\beta!}\Sigma'_{\beta}(\nu) = \frac{1}{\beta!}\Sigma'\frac{\nu!}{\nu_1!\,\nu_2!\ldots\nu_{\beta}!}. \tag{23}$$

IV. Proof of normality of p' for $n \to \infty$

12. Any set of different pairs of elements belonging to the population will be briefly referred to as a *system* (two pairs being different if they have no more than one element in common).

If we represent the elements of a system by points in a plane and the pairs of elements by lines joining the points, we have a *pattern* corresponding to the given system. Two systems will be said to have the same pattern if there exists a one-to-one correspondence between the elements of both systems such that if two elements of one system form a pair, the two corresponding elements of the other system also form a pair. Thus the only thing relevant in a pattern is the lines connecting the points, the position of the points having no significance.

A pattern will be called *simple* if one can pass from any point of the pattern to any other one along lines belonging to the pattern. A *composite* pattern is a pattern consisting of more than one simple pattern.

If the elements of a system (or the points of the corresponding pattern) are denoted by different letters A, B, C, ..., each pair of the system can be represented by a pair of letters. All systems of one pair have the same pattern (AB). There are two patterns of two pairs, one simple and containing three points (AB, BC) and one composite and containing four points (AB, CD). There are five patterns of three pairs, three simple $(AB, BC, CA;\ AB, BC, CD;\ AB, AC, AD)$, one consisting of two different simple patterns (AB, CD, DE) and one consisting of three equal simple patterns (AB, CD, EF).

13. If a *simple* pattern consists of a_j points and b_j pairs,

$$a_j \leqslant b_j + 1. \tag{24}$$

For this is true for $b_j = 1$, and by adding one pair to a simple pattern, at most one point is added if the new pattern is to be simple again.

Denote the different simple patterns by $S_1, S_2, \ldots, S_j, \ldots$, where S_1 stands for the one-pair pattern and S_2 for the two-pairs pattern (AB, BC), all S_j with $j \geqslant 3$ consisting of three or more pairs. Let a_j be the number of points and b_j the number of pairs in S_j. Then $a_1 = 2$, $a_2 = 3$,

$$b_1 = 1, \quad b_2 = 2, \quad b_j \geqslant 3 \quad \text{if} \quad j \geqslant 3. \tag{25}$$

Consider a pattern P composed of γ_1 simple patterns S_1, γ_2 simple patterns S_2, etc., and containing a points and b pairs. Then, writing symbolically

$$P = \Sigma \gamma_j S_j,$$

we have

$$a = \Sigma \gamma_j a_j, \quad b = \Sigma \gamma_j b_j.$$

In virtue of (24),

$$3b - 2a = \Sigma \gamma_j (3b_j - 2a_j) \geqslant \Sigma \gamma_j (b_j - 2),$$

and from (25)

$$3b - 2a \geqslant -\gamma_1, \tag{26}$$

the sign of equality holding if, and only if, pattern P contains no other simple patterns than S_1 and S_2.

14. $(p')^\mu$ is the probability that μ pairs of elements drawn from the sample, replacing each pair after drawing, are all concordant. We may write

$$(p')^\mu = \Sigma A_i w_i',$$

where A_i is the probability that μ pairs are drawn from the sample in such a way that the system of different pairs among them has the pattern P_i, and if a_i is the number of points in P_i, w_i' is the probability that if a_i elements are drawn from the sample without replacement

and paired according to pattern P_i, all pairs of P_i are concordant. The summation is extended over all patterns P_i with no more than μ pairs.

Since the probabilities w_i' are of the type for which formula (9) is applicable, we have

$$E(p')^\mu = \Sigma A_i w_i, \tag{27}$$

where, as usual, w_i is the population probability corresponding to the sample probability w_i'.

15. Consider a term Aw in (27) corresponding to the pattern

$$P = \Sigma \gamma_j S_j$$

with $\gamma = \Sigma \gamma_j$ simple patterns, $a = \Sigma \gamma_j a_j$ points and $b = \Sigma \gamma_j b_j$ pairs.

Let

$$\overline{P} = \sum_{j \geqslant 2} \gamma_j S_j$$

be the pattern obtained from P by excluding the single-pair patterns S_1. Then

$$\overline{\gamma} = \gamma - \gamma_1, \quad \overline{a} = a - 2\gamma_1 \quad \text{and} \quad \overline{b} = b - \gamma_1 \tag{28}$$

are the numbers of simple patterns, points and pairs in $\overline{P}$.

We have

$$w = p^{\gamma_1} v, \tag{29}$$

where v is independent of p and γ_1, only depending on the pattern $\overline{P}$.

16. The probability A will be studied, in the first place, as a function of n and γ_1, while its dependence on $\overline{P}$ will be considered later and only in a special case. It must be borne in mind that, by (28), γ, a and b also depend on γ_1.

Let $Q_1, Q_2, \ldots, Q_b$ be the pairs of pattern P numbered in some definite order. Suppose pair Q_β appears μ_β times ($\beta = 1, \ldots, b$). Then

$$\mu_1 + \ldots + \mu_b = \mu, \quad \mu_\beta \geqslant 1 \quad (\beta = 1, \ldots, b).$$

Let $R_1, R_2, \ldots, R_\mu$ be the total set of the pairs drawn, numbered independently of the order in which they appear. Then μ_1 R's are equal to Q_1, μ_2 R's are equal to Q_2, etc.

Let B be the probability that among μ pairs drawn from the sample, replacing each pair after drawing, b pairs are different and arranged according to pattern P, pair Q_β appearing μ_β times ($\beta = 1, \ldots, b$) and the μ pairs being drawn in a definite order, say $R_1, R_2, \ldots, R_\mu$.

Suppose, R_1 is a Q_1. Since any pair drawn may be taken as Q_1 (only the relative position of the pairs being relevant), the probability first to draw R_1 is 1. The probability that the second pair drawn is R_2 depends on whether R_2 has no, one or both elements in common with R_1. In the first case, it is $\binom{n-2}{2} \Big/ \binom{n}{2}$, in the second case $2(n-2) \Big/ \binom{n}{2}$ (the factor 2 arising from the fact that each of the two elements of R_1 can be the element common with R_2), and in the third case, $1 \Big/ \binom{n}{2}$.

In general, if the first λ pairs drawn are $R_1, \ldots, R_\lambda$, and if they form a pattern P' containing α different elements, the probability that the $(\lambda+1)$th pair drawn is $R_{\lambda+1}$ depends on whether $R_{\lambda+1}$ has no, one or both elements in common with P'. In the first case it is $\binom{n-\alpha}{2} \Big/ \binom{n}{2}$, in the second case, $c'(n-\alpha) \Big/ \binom{n}{2}$, and in the third case, $c'' \Big/ \binom{n}{2}$, where c' and c'' are independent of n. If, in the last case, $R_{\lambda+1}$ is equal to one of the preceding R's, $c'' = 1$.

B is the product of all μ such probabilities, and it is seen from the above consideration that it is of the form

$$B = C(n-2)^{[a-2]}\binom{n}{2}^{-\mu+1},$$

where C is independent of n.

We also see that a pair which has already appeared before makes no contribution to C. Hence, C only depends on the different pairs of pattern P, and is independent of the numbers μ_β.

The above reflexion further shows that for any simple pattern contained in P, the pair drawn first, having no elements in common with the preceding pairs, contributes to C the factor $\frac{1}{2}$, except for the first pair, R_1, which yields the factor 1. Thus, C contains the factor $2^{-\gamma+1} = 2^{-\gamma_1-\bar{\gamma}+1}$, and $2^{-\gamma_1}$ is obviously the sole contribution to C from the γ_1 single pairs (pattern S_1) contained in P. Hence

$$B = 2^{-\gamma_1}C'(n-2)^{[a-2]}\binom{n}{2}^{-\mu+1},$$

where C' is independent of n and γ_1, and also independent of the order in which the γ_1 single pairs are drawn.

A, the probability that μ pairs drawn form pattern P, irrespective of the order in which they appear, depends on n in the same way as B. As a function of γ_1, A contains, besides $2^{-\gamma_1}$, the factor $1/\gamma_1!$ owing to the fact that the γ_1 single pairs are interchangeable. Further it contains the factor $\Sigma'_b(\mu)$ which indicates the number of ways of allocating μ objects on b places so that no place remains empty. In virtue of (23) we have

$$\Sigma'_b(\mu) = b!\, d_{\mu,\mu-b}.$$

Thus, A is of the form

$$A = D^{(\mu)}_{\gamma_1}(n-2)^{[2\gamma_1+\bar{a}-2]}\binom{n}{2}^{-\mu+1}, \tag{30}$$

where

$$D^{(\mu)}_{\gamma_1} = 2^{-\gamma_1}\frac{(\gamma_1+\bar{b})!}{\gamma_1!}d_{\mu,\mu-\gamma_1-\bar{b}}D' \tag{31}$$

and D' is independent of both n and γ_1 and only depends on the pattern $\overline{P}$ containing no S_1.

Inserting (29) and (30) in (27), we have

$$\binom{n}{2}^{\mu-1}E(p')^\mu = \Sigma D^{(\mu)}_{\gamma_1}p^{\gamma_1}v(n-2)^{[2\gamma_1+\bar{a}-2]}, \tag{32}$$

the summation taking place over all patterns with no more than μ pairs. (32) also holds for $\mu = 0$ if by a 'pattern of 0 pairs' we understand the case $\gamma_1 = \gamma_2 = \ldots = 0$ and take (31) as definition of $D^{(0)}_{\gamma_1}$ with suitably chosen D'. $\rho^{[-\delta]}$ with $\delta > 0$ is defined by

$$\rho^{[-\delta]}(\rho+\delta)^{[\delta]} = 1.$$

17. If

$$\mu_\nu(p') = E(p'-p)^\nu,$$

we have

$$\binom{n}{2}^{\nu-1}\mu_\nu(p') = \sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\binom{n}{2}^\delta p^\delta\binom{n}{2}^{\nu-\delta-1}E(p')^{\nu-\delta}.$$

Applying (32) with

$$\mu = \nu-\delta, \quad \gamma_1 = \kappa-\delta, \tag{33}$$

we have for the coefficient of $p^\kappa v$ in $\binom{n}{2}^{\nu-1}\mu_\nu(p')$

$$\sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\binom{n}{2}^\delta D^{(\nu-\delta)}_{\kappa-\delta}(n-2)^{[\bar{a}+2\kappa-2\delta-2]}$$

$$= \sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\frac{1}{2^\delta}\sum_{\rho=0}^{\delta}(-1)^\rho\binom{\delta}{\rho}n^{2\delta-\rho}D^{(\nu-\delta)}_{\kappa-\delta}(n-2)^{[\bar{a}+2\kappa-2\delta-2]}.$$

Inserting here, in accordance with (15),

$$n^{2\delta-\rho} = \sum_{\sigma=0}^{2\delta-\rho} d_{2\delta-\rho,\sigma}^{(\bar{a}+2\kappa-2\delta)}\,(n-\bar{a}-2\kappa+2\delta)^{[2\delta-\rho-\sigma]},$$

we have $$\sum_{\delta=0}^{\nu}(-1)^{\delta}\binom{\nu}{\delta}\frac{1}{2^{\delta}}\sum_{\rho=0}^{\delta}(-1)^{\rho}\binom{\delta}{\rho}D_{\kappa-\delta}^{(\nu-\delta)}\sum_{\sigma=0}^{2\delta-\rho} d_{2\delta-\rho,\sigma}^{(\bar{a}+2\kappa-2\delta)}\,(n-2)^{(\bar{a}+2\kappa-2-\rho-\sigma]}.$$

Putting $\alpha = \bar{a}+2\kappa-2-\rho-\sigma$, we have for the coefficient $K_{\kappa}^{(\nu,\alpha)}$ of $p^{\kappa}\nu(n-2)^{(\alpha]}$ in $\binom{n}{2}^{\nu-1}\mu_{\nu}(p')$

$$K_{\kappa}^{(\nu,\alpha)} = \sum_{\delta=0}^{\nu}(-1)^{\delta}\binom{\nu}{\delta}a_{\kappa}^{(\nu,\alpha)}(\delta), \tag{34}$$

where $$a_{\kappa}^{(\nu,\alpha)}(\delta) = \frac{1}{2^{\delta}}D_{\kappa-\delta}^{(\nu-\delta)}\sum_{\rho=0}^{\delta}(-1)^{\rho}\binom{\delta}{\rho}d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)}. \tag{35}$$

Since $d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)} = 0$ if $\bar{a}+2\kappa-\alpha-\rho-2<0$, the upper limit of ρ in the summation may be taken as $\bar{a}+2\kappa-\alpha-2$, which is independent of δ. We have then, in virtue of (31),

$$a_{\kappa}^{(\nu,\alpha)}(\delta) = \frac{1}{2^{\kappa}}(\bar{b}+\kappa-\delta)^{[\bar{b}]}d_{\nu-\delta,\nu-\bar{b}-\kappa}D'\sum_{\rho=0}^{\bar{a}+2\kappa-\alpha-2}(-1)^{\rho}\binom{\delta}{\rho}d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)}, \tag{36}$$

where D' is independent of δ.

In virtue of (I) and (II), para. 10, $a_{\kappa}^{(\nu,\alpha)}(\delta)$ is a polynomial in δ. The degree of the $(\rho+1)$th term in the sum in (36) is $\rho+2(\bar{a}+2\kappa-\alpha-\rho-2)$, which is highest for $\rho=0$. Hence, the degree d of $a_{\kappa}^{(\nu,\alpha)}(\delta)$ is

$$d = \bar{b}+2(\nu-\bar{b}-\kappa)+2(\bar{a}+2\kappa-\alpha-2) = 2(\nu+\bar{a}+\kappa-\alpha-2)-\bar{b}. \tag{37}$$

Now, according to (34) and (14), $K_{\kappa}^{(\nu,\alpha)}=0$ if $d<\nu$, or, in virtue of (37),

$$K_{\kappa}^{(\nu,\alpha)} = 0 \quad \text{if} \quad 2\alpha > \nu-4+2\bar{a}-\bar{b}+2\kappa. \tag{38}$$

Applying (26) for pattern $\bar{P}$, we have, since $\gamma_1=0$, $2\bar{a}\leqslant 3\bar{b}$, and consequently

$$2\bar{a}-\bar{b}+2\kappa \leqslant 2(\bar{b}+\kappa),$$

the sign of equality holding if and only if pattern P contains no other simple patterns than S_1 and S_2.

Remembering that, according to para. 16, $b\leqslant\mu$, we have in virtue of (33)

$$\bar{b}+\kappa = \bar{b}+\gamma_1+\delta = b+\delta \leqslant \mu+\delta = \nu. \tag{39}$$

Thus in any case $$2\bar{a}-\bar{b}+2\kappa \leqslant 2\nu$$

and, in virtue of (38), $$K_{\kappa}^{(\nu,\alpha)} = 0 \quad \text{if} \quad \alpha \geqslant \tfrac{3}{2}\nu-1. \tag{40}$$

If P contains at least one simple pattern with more than two pairs, we even have

$$2\bar{a}-\bar{b}+2\kappa < 2\nu,$$

and consequently $$K_{\kappa}^{(\nu,\alpha)} = 0 \quad \text{if} \quad \alpha \geqslant \tfrac{3}{2}\nu-2. \tag{41}$$

From (40) it appears that the degree in n of $\binom{n}{2}^{\nu-1}\mu_{\nu}(p')$ is

$$\leqslant 3h-2 = \tfrac{3}{2}\nu-2 \quad \text{if} \quad \nu = 2h,$$

$$\leqslant 3h-1 = \tfrac{3}{2}\nu-\tfrac{5}{2} \quad \text{if} \quad \nu = 2h+1.$$

Thus, in $\mu_{2h+1}(p')$, if expanded in powers of n, the degree of the highest term is

$$\leqslant 3h-1-4h = -h-1.$$

In $\mu_3(p')$, in virtue of (13), the degree of the highest term is -1, provided that $k-p^2 \neq 0$. Hence, the degree of

$$\alpha_{2h+1}(p') = \frac{\mu_{2h+1}(p')}{\mu_2^{\frac{1}{2}(2h+1)}(p')}$$

is $\leqslant -h-1+h+\frac{1}{2} = -\frac{1}{2}$. It follows that

$$\alpha_{2h+1}(p') \underset{n}{\to} 0 \quad \text{if} \quad k-p^2 > 0. \tag{42}$$

($k-p^2 < 0$ is impossible since in this case var (p') would become < 0 for large n.)

18. As we have seen, we may write

$$\binom{n}{2}^{2h-1} \mu_{2h}(p') = R_h (n-2)^{[3h-2]} + R_h' (n-2)^{[3h-3]} + \ldots.$$

Then it follows from (41) that R_h only contains terms depending on patterns S_1 and S_2, that is, R_h is of the form

$$R_h = \sum_{\kappa} \sum_{\lambda} K_{\kappa,\lambda}^{(2h,3h-2)} p^{\kappa} k^{\lambda}. \tag{43}$$

The only terms in $\binom{n}{2}^{\mu-1} E(p')^{\mu}$ which can contribute to this sum are of the form

$$D_{\gamma_1}^{(\mu)} p^{\gamma_1} k^{\lambda} (n-2)^{[2\gamma_1+3\lambda-2]}.$$

The pattern corresponding to such a term is

$$P = \gamma_1 S_1 + \lambda S_2.$$

Remembering the considerations in para. 16, we see that in each S_2 the pair drawn first contributes to C the factor $\frac{1}{2}$ (except if it is R_1), while the first drawing of the other pair yields the factor 2. Hence, the S_2's make no contribution to C, and we have

$$C = 2^{-\gamma_1+1}.$$

The contribution of the patterns S_2 to A is twofold: since in each S_2 the two pairs may be interchanged, this gives the factor $(1/2!)^{\lambda}$; and since the λ patterns S_2 may be interchanged, we have the factor $1/\lambda!$. Thus

$$D_{\gamma_1}^{(\mu)} = 2^{-\gamma_1-\lambda+1} \frac{(\gamma_1+2\lambda)!}{\gamma_1!\,\lambda!} d_{\mu,\,\mu-\gamma_1-2\lambda}$$

and, in virtue of (31), since $\bar{b} = 2\lambda$,

$$D' = \frac{1}{2^{\lambda-1}\lambda!}. \tag{44}$$

Inserting in (38) $\quad \nu = 2h, \quad \alpha = 3h-2, \quad \bar{a} = 3\lambda, \quad \bar{b} = 2\lambda, \tag{45}$

we see that $K_{\kappa,\lambda}^{(2h,3h-2)} \neq 0$ is possible only if

$$\kappa + 2\lambda \geqslant 2h.$$

On the other hand, from (39), $\quad \kappa + 2\lambda \leqslant 2h.$

Hence, $\quad \kappa = 2h - 2\lambda. \tag{46}$

Inserting this in (43), we have

$$R_h = \sum_{\lambda=0}^{h} K_{2(h-\lambda),\lambda}^{(2h,3h-2)} p^{2(h-\lambda)} k^{\lambda}, \tag{47}$$

where

$$K_{2(h-\lambda),\lambda}^{(2h,3h-2)} = \sum_{\delta=0}^{2h} (-1)^{\delta} \binom{2h}{\delta} a_{2(h-\lambda)}^{(2h,3h-2)}(\delta).$$

According to (37) in connexion with (45), (46), the degree in δ of $a_{2\lambda,\lambda}^{(2\lambda,2\lambda-2\lambda)}(\delta)$ is 2λ. The highest term, $a_0\delta^{2\lambda}$, is contained in the term corresponding to $\rho = 0$ in (36). Inserting in (36) the values from (44), (45) and (46) and putting in the sum $\rho = 0$, we have

$$2^{-2\lambda+\lambda+1}\frac{(2h-\delta)^{[2\lambda]}}{\lambda!}d_{2\lambda-\delta,0}a_{2\lambda,h-\lambda}^{(4h-\lambda-2\delta)}.$$

Thus, in virtue of (16) and (II), para. 10,

$$a_0 = 2^{-2\lambda+\lambda+1}\frac{1}{\lambda!}(-1)^{h-\lambda}\frac{2^{2h-2\lambda}}{2^{h-\lambda}(h-\lambda)!} = \frac{(-1)^{h-\lambda}}{2^{h-1}h!}\binom{h}{\lambda}.$$

According to (14),
$$K_{2(h-\lambda),\lambda}^{(2\lambda,2h-2\lambda)} = \frac{(-1)^{h-\lambda}(2h)!}{2^{h-1}h!}\binom{h}{\lambda}.$$

Inserting this in (47) we have

$$R_h = \frac{(2h)!}{2^{h-1}h!}\sum_{\lambda=0}^{h}(-1)^{h-\lambda}\binom{h}{\lambda}p^{2(h-\lambda)}k^{\lambda} = \frac{(2h)!}{2^{h-1}h!}(k-p^2)^h.$$

The highest term of $\mu_{2h}(p')$ is thus

$$2^h\frac{(2h)!}{h!}(k-p^2)^h n^{-h},$$

that of $\mu_2(p')$ is $4(k-p^2)n^{-1}$, and hence

$$\alpha_{2h}(p') = \frac{\mu_{2h}(p')}{\mu_2^h(p')} \underset{n}{\to} \frac{(2h)!}{2^h h!} \quad \text{if} \quad k-p^2 > 0. \tag{48}$$

From (42) and (48) it follows according to the Second Limit Theorem that the distribution of p' tends to normality as $n \to \infty$, provided that the marginal distributions are continuous and $k-p^2 > 0$.

The condition $k-p^2 > 0$ is fulfilled if the population is distributed normally. For, comparing Esscher's formula (5) with (13), we find, since $\operatorname{var}(r) = 4\operatorname{var}(p')$,

$$k-p^2 = \frac{1}{36} - \left\{\frac{1}{\pi}\sin^{-1}\left(\frac{r}{2}\right)\right\}^2.$$

The right-hand side is positive if $|r| < 1$.

V. THE VARIANCE OF p' IN THE CASE OF A FINITE POPULATION

19. Consider a sample of n drawn from a finite population of N in which all values of x and all values of y are different. For the sample probabilities p', k', $(p^2)'$, ... we write now

$$p^{(n)}, \quad k^{(n)}, \quad (p^2)^{(n)}, \quad \ldots,$$

and for the corresponding population probabilities

$$p^{(N)}, \quad k^{(N)}, \quad (p^2)^{(N)}, \quad \ldots.$$

Equation (9) remains valid and may be written as follows:

$$Ew^{(n)} = w^{(N)}. \tag{49}$$

In particular,
$$Ep^{(n)} = p^{(N)}.$$

The essential difference between this case and the case $N = \infty$ considered above is that the composite probabilities such as $(p^2)^{(N)}$ or $(pk)^{(N)}$ are not equal to $(p^{(N)})^2$ or $p^{(N)}k^{(N)}$. For

instance, $(p^{(N)})^2$ is evidently the same function of $p^{(N)}$, $k^{(N)}$, $(p^2)^{(N)}$ and N as $(p')^2$ is of p', k', $(p^2)'$ and n. Thus we have, replacing n by N in (10),

$$(p^{(N)})^2 = \frac{2}{N^{[2]}}p^{(N)} + \frac{4(N-2)}{N^{[2]}}k^{(N)} + \frac{(N-2)^{[2]}}{N^{[2]}}(p^2)^{(N)}, \tag{50}$$

and hence

$$(p^2)^{(N)} = \frac{N^{[2]}}{(N-2)^{[2]}}(p^{(N)})^2 - \frac{4}{N-3}k^{(N)} - \frac{2}{(N-2)^{[2]}}p^{(N)}. \tag{51}$$

On the other hand, from (10) and (49)

$$E(p^{(n)})^2 = \frac{2}{n^{[2]}}p^{(N)} + \frac{4(n-2)}{n^{[2]}}k^{(N)} + \frac{(n-2)^{[2]}}{n^{[2]}}(p^2)^{(N)}, \tag{52}$$

which is the equivalent of equation (11).

On subtracting (50) from (52) we find

$$\operatorname{var}(p^{(n)}) = \frac{2(N-n)}{n(n-1)N(N-1)}\{(N+n-1)p^{(N)} + 2[Nn - 2(N+n-1)]k^{(N)} - [2Nn - 3(N+n-1)](p^2)^{(N)}\}. \tag{53}$$

Substituting for $(p^2)^{(N)}$, the expression in (51), we obtain

$$\operatorname{var}(p^{(n)}) = \frac{2(N-n)}{n(n-1)(N-2)(N-3)}\{(N+n-5)p^{(N)}(1-p^{(N)}) + 2(n-2)(N-2)(k^{(N)} - p^{(N)2})\}, \tag{54}$$

or

$$\binom{n}{2}\operatorname{var}(p^{(n)}) = \left(1 - \frac{(n-2)(n-3)}{(N-2)(N-3)}\right)p^{(N)}(1-p^{(N)}) + 2(n-2)\left(1 - \frac{n-3}{N-3}\right)(k^{(N)} - p^{(N)2}). \tag{55}$$

For $N \to \infty$, (55) becomes the same as (13)

VI. A SAMPLE ESTIMATE OF var (p')

20. In the case of an infinite population, let

$$\binom{n}{2}\operatorname{var}'(p') = p' + 2(n-2)k' - (2n-3)(p^2)'. \tag{56}$$

Then, in virtue of (9) and (12),

$$E\operatorname{var}'(p') = \operatorname{var}(p').$$

On inserting in (56) for $(p^2)'$ the expression obtained from (10), we find

$$\binom{n-2}{2}\operatorname{var}'(p') = p'(1-p') + 2(n-2)(k' - (p')^2),$$

or

$$\operatorname{var}'(p') = \frac{2}{(n-2)(n-3)}p'q' + \frac{4}{n-3}(k' - (p')^2). \tag{57}$$

By analogy,

$$\operatorname{var}'(q') = \frac{2}{(n-2)(n-3)}p'q' + \frac{4}{n-3}(l' - (q')^2), \tag{58}$$

where l' is the probability that among three members A, B, C drawn from the sample without replacement, one, say A, is discordant with the other two.

In the case of a finite population of the type considered in para. 19, we define in a similar way a statistic $\operatorname{var}^{(n)}(p^{(n)})$ such that

$$E\operatorname{var}^{(n)}(p^{(n)}) = \operatorname{var}(p^{(n)}).$$

We find

$$\operatorname{var}^{(n)}(p^{(n)}) = \frac{2(N-n)}{N(N-1)(n-2)(n-3)}\{(N+n-5)\,p^{(n)}(1-p^{(n)}) + 2(n-2)(N-2)(k^{(n)}-p^{(n)2})\} \tag{59}$$

A comparison between (59) and (54) shows that $\operatorname{var}^{(n)}(p^{(n)})$ is obtained from $\operatorname{var}(p^{(n)})$ by interchanging n and N and taking the opposite sign.

21. Let g_ν and h_ν be the numbers of sample members concordant and discordant with $A_\nu = (x_\nu, y_\nu)$ $(\nu = 1, \ldots, n)$. The probability of drawing first the member A_ν, and then, without replacing it, a member concordant with A_ν is $\frac{1}{n}\frac{g_\nu}{n-1}$. The probability of drawing, without replacement, first A_ν and then two other members concordant with A_ν is $\frac{1}{n}\frac{g_\nu(g_\nu-1)}{(n-1)(n-2)}$. Hence

$$p' = \frac{\Sigma g_\nu}{n(n-1)}, \quad k' = \frac{\Sigma g_\nu^2 - \Sigma g_\nu}{n(n-1)(n-2)}. \tag{60}$$

Similarly,

$$q' = \frac{\Sigma h_\nu}{n(n-1)}, \quad l' = \frac{\Sigma h_\nu^2 - \Sigma h_\nu}{n(n-1)(n-2)}. \tag{61}$$

If only the value of p' or q' is required, the use of (6) may be more expedient than that of (60) or (61). If, however, the variance, and hence k' or l', is wanted, the calculation by means of the numbers g_ν and h_ν (whose sums are twice the numbers of inversions I, J) according to (60) or (61) is to be preferred.

If $p' > \frac{1}{2}$, it is more convenient to calculate q' and l' from (61); if $p' < \frac{1}{2}$, the calculation of p' and k' by (60) is more rapid. In many cases one can see directly from the given data whether the concordant or the discordant pairs prevail, before actually calculating p' or q'.

Since $p' + q' = 1$, we have $\operatorname{var}(p') = \operatorname{var}(q')$,

and also, in the case of a finite population,

$$\operatorname{var}(p^{(n)}) = \operatorname{var}(q^{(n)}).$$

If we write down the equation for $\operatorname{var}(q^{(n)})$ analogous to (55) and subtract it from (55), we have

$$k^{(N)} - p^{(N)2} - l^{(N)} + q^{(N)2} = 0,$$

or

$$k^{(N)} - l^{(N)} = p^{(N)2} - q^{(N)2} = p^{(N)} - q^{(N)}.$$

Substituting n for N, we have

$$k' - l' = p' - q' = \tau.$$

Comparing this with (57) and (58) we see that

$$\operatorname{var}'(p') = \operatorname{var}'(q').$$

REFERENCES

ESSCHER, F. (1924). On a method of determining correlation from the ranks of the variates. *Skand. Aktuar.* **7**, 201–19.

JORDAN, CH. (1927). *Statistique mathématique.* Paris: Gauthiers-Villars.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81–93.

LINDEBERG, J. W. (1926). Ueber die Korrelation. *Den VI skandinaviske Matematikerkongres i København*, 31 August–4 September 1925, pp. 437–46. København: J. Gjellerup.

ADDENDUM

On p. 184 above, I quoted J. W. Lindeberg as having given the formula for the variance of the probability of concordance p' without proof. I was not aware then that a proof of this formula, as well as that of the corresponding expression for a finite population (equation (54) of my paper), is contained in another paper by Lindeberg, 'Some remarks on the mean error of the percentage of correlation,' *Nordic Statistical Journal*, **1**, 137–41 (1929).

THE SIGNIFICANCE OF RANK CORRELATIONS WHERE PARENTAL CORRELATION EXISTS

By H. E. DANIELS (*Wool Industries Research Association*)
and M. G. KENDALL

1. All the known tests of significance of rank correlation coefficients are based on distributions from a population in which each possible ranking occurs equally frequently, i.e. the null case where no parental correlation exists. We may then say of any particular coefficient whether it is significant in the sense that it cannot have arisen with any acceptable probability from an uncorrelated population. No tests are known in the case where parental correlation exists, and we have not seen the point discussed except in reference to the replacement of rank correlations by grade or product-moment correlations. Thus, for example, if two rank correlation coefficients are both found to be significant there has hitherto been no exact method of deciding whether their difference is significant. In this paper we consider the problem of determining confidence intervals for a rank correlation when the parent is correlated and develop a test of significance for the difference of two correlations.

2. In testing an ordinary product-moment correlation the problem is enormously simplified by the assumption that the population is normal, or the further assumption that normal theory holds good even when the parent deviates only moderately from normality. Apart from means and variances the population is then completely specified by the single parent parameter ρ and, as is well known, the sample distribution of the estimator depends only on ρ and the sample number n.

In ranking theory this position no longer obtains. No assumption can in general be made about the form of the parent distribution and, in particular, the parent correlation does not completely specify the problem. The usual type of variate theory cannot, therefore, be expected to meet the requirements.

3. A satisfactory approach to the problem can, however, be made if the rank correlation is measured by the coefficient known as τ (Kendall, 1943, chap. 16). We shall then show that, for large samples at any rate, the problem admits of a solution.

Let the population consist of N members. They may be imagined as laid out in the natural order $1, 2, \ldots, N$ according to the first variate. The rankings according to the second variate are then some permutation of the numbers 1 to N, and this second array of ranks is all we need write down in particular cases. It determines the rank correlation τ. Now suppose we choose a sample of n in one of the $\binom{N}{n}$ possible ways. This sample will, so far as the first variate is concerned, be in the natural order, and the ranks according to the second variate permit of the calculation of a sample correlation t. For all possible samples and any given arrangement of the parent members there will be a distribution of $\binom{N}{n}$ values of t.

4. The sample value of t is an unbiased estimator of τ; that is to say, the mean value of t in all possible samples is τ. For consider the $\binom{N}{n}$ samples of n. Any particular pair of members will occur in $\binom{N-2}{n-2}$ samples, that is, all pairs occur equally frequently in the totality of all samples. In calculating t we assign to any pair $+1$ if its members are in the right order and -1 in the contrary case. Thus the total of the score for all samples is $\binom{N-2}{n-2}$ times the score for the population. To obtain t we divide the score for any sample by $\frac{1}{2}n(n-1)$, and to obtain τ we divide the population score by $\frac{1}{2}N(N-1)$. Hence if Σ is the score for the population, the mean value (expectation) of t is

$$E(t) = \frac{\binom{N-2}{n-2}\Sigma}{\frac{1}{2}n(n-1)\binom{N}{n}} = \frac{\Sigma}{\frac{1}{2}N(N-1)} = \tau. \tag{4·1}$$

5. Unfortunately, it is not true that higher moments of t depend only on τ. A single example will illustrate the point. Consider the ranking of 9:

5 2 3 1 6 7 8 9 4.

If the $84 = \binom{9}{3}$ possible samples of three are written down and t evaluated for each, the distribution of S (the number of positive pairs) is found to be as follows:

Values of S	Frequency
0	2
1	15
2	34
3	33
Total	84

The mean of this distribution is $182/84 = 13/6$, and since

$$t = \frac{2S}{\frac{1}{2}n(n-1)} - 1,$$

the mean value of t is $(26/18) - 1 = 0{\cdot}44$. The value of S for the parent ranking is 26 and hence $\tau = (52/36) - 1 = 0{\cdot}44$, verifying equation (4·1). The ranking

1 2 5 9 3 6 7 8 4

also has $\tau = 0{\cdot}44$, but the distribution of S in samples of three is now:

Values of S	Frequency
0	3
1	16
2	29
3	36
Total	84

The second moment of this distribution is 5·429, against 5·333 for the first distribution, the variances being 0·734 against 0·639.

6. Thus for any parent with given τ there is in general more than one sampling distribution of t according to the arrangement of the parent ranks. In short, as mentioned above, the parameter τ does not completely specify the sampling distribution and in asking the question: What is the standard error of t? we are seeking for an answer which does not exist.

It will be shown, however, that for any given parent ranking the distribution of t tends to normality with increasing n. The sampling properties of t can therefore be specified to a first approximation by its first and second moments only, when the samples are not too small. Further, it will be proved that for given τ the variance of t cannot exceed a certain function of τ and n whatever the parent ranking. From a knowledge of t and n only, it is thus possible to set outer bounds to confidence intervals for τ provided n is large enough for the normal approximation to hold. The limits obtained in this way are sometimes rather wide, and an alternative procedure is to estimate the true variance of t directly from the sample itself according to a formula given below. This avoids the loss of efficiency consequent on using an upper limit to the variance, but it is not known how large a sample is required for the error of estimation to be tolerable.

7. The development of the theory is facilitated if we introduce at the present stage a notation similar to that used by Daniels (1944). The ith and jth ranks corresponding to the second variate are together assigned a score a_{ij} which takes the value $+1$ if the members are in the correct order, -1 if in the wrong order, and a_{ii} is defined to be zero. The ranks for the first variate are similarly assigned scores b_{ij}, but as the members have been taken in the correct order for this variate, the scores are simply $b_{ij} = \pm 1$, $i \lesssim j$; $b_{ii} = 0$. Next we define $c_{ij} = a_{ij}b_{ij}$, so that $c_{ij} = \pm 1$, according to whether the ranks for the two variates agree or differ in order, and $c_{ii} = 0$. In this notation

$$\tau = c/N(N-1),$$

where $c = \Sigma c_{ij}$, i and j both being summed from 1 to N.

When the sample of n pairs is selected at random from the parent N and its coefficient t is calculated, the values of c_{ij} for the members of the sample remain the same as in the population. This fact makes τ much more suitable for the present problem than the Spearman coefficient ρ whose associated scores do not possess the same property. The sample rank correlation is then

$$t = c^{(n)}/n(n-1),$$

where $c^{(n)} = \Sigma^{(n)}c_{ij}$ and $\Sigma^{(n)}$ denotes summation only over those values of i and j occurring in the sample.

8. It has already been proved that $E(t) = \tau$. To find the variance of t we require $E(t^2)$, so consider

$$\sum_n [c^{(n)}]^2 = \sum_n \Sigma^{(n)} c_{ij} c_{kl},$$

$\sum_n$ denoting summation over all selections of the sample of n from the finite parent population of N members. Let us enumerate the number of ways in which $c_{ij}c_{kl}$ and similar products with 'tied' suffixes, such as $c_{ij}c_{il}$, occur in the sum.

(i) When i, j, k, l are all different the term $c_{ij}c_{kl}$ may occur with $\binom{N-4}{n-4}$ selections of the remaining members of the sample and the contribution of such terms to $\sum_n$ is $\binom{N-4}{n-4}\Sigma' c_{ij}c_{kl}$, Σ' meaning summation over all unequal values of i, j, k, l from 1 to N.

(ii) The term $c_{ij}c_{il}$ similarly occurs in $\binom{N-3}{n-3}$ ways and there are four ways of tying one suffix, each of which gives the same contribution to Σ since c_{ij} is symmetrical. The total contribution of such terms to $\sum_n$ is therefore $4\binom{N-3}{n-3}\Sigma' c_{ij}c_{il}$.

(iii) Terms like $c_{ij}c_{ij}$ similarly contribute $2\binom{N-2}{n-2}\Sigma' c_{ij}c_{ij}$ to $\sum_u$, and all other terms are zero since $c_{ii}=0$. Hence

$$\sum_n [c^{(n)}]^2 = \binom{N-4}{n-4}\Sigma' c_{ij}c_{kl} + 4\binom{N-3}{n-3}\Sigma' c_{ij}c_{ik} + 2\binom{N-2}{n-2}\Sigma' c_{ij}c_{ij}.$$

Expressing the Σ''s in terms of the corresponding Σ's and dividing out by $\binom{N}{n}$ we obtain

$$E[c^{(n)}]^2 = \frac{n^{(4)}}{N^{(4)}}(\Sigma c_{ij}c_{kl} - 4\Sigma c_{ij}c_{il} + 2\Sigma c_{ij}c_{ij}) + \frac{4n^{(3)}}{N^{(3)}}(\Sigma c_{ij}c_{il} - \Sigma c_{ij}c_{ij}) + \frac{2n^{(2)}}{N^{(2)}}\Sigma c_{ij}c_{ij},$$

where $n^{(r)} = n(n-1)\dots(n-r+1)$. Since $\Sigma c_{ij}c_{ij} = N(N-1)$ and $\Sigma c_{ij}c_{kl} = c^2$, the variance of t for given τ and n is seen to depend on the value of $\Sigma c_{ij}c_{ik} = \Sigma c_i^2$, where $c_i = \sum_{j=1}^{N} c_{ij}$.

Let N become large. The quantities c and Σc_i^2 are respectively $O(N^2)$ and $O(N^3)$, so if we introduce $\tau_i = c_i/N$ the value of $E(t^2)$ for large N becomes

$$E(t^2) \sim \frac{(n-2)(n-3)}{n(n-1)}\tau^2 + \frac{4(n-2)}{n(n-1)}\frac{\Sigma\tau_i^2}{N} + \frac{2}{n(n-1)},$$

and hence in the limit the variance of t is

$$\operatorname{var} t = \frac{4(n-2)}{n(n-1)}\operatorname{var}\tau_i + \frac{2}{n(n-1)}(1-\tau^2). \tag{8·1}$$

9. The variance of t satisfies the inequality

$$\operatorname{var} t \leqslant \frac{2}{n}(1-\tau^2), \tag{9·1}$$

whatever the parent ranking. Moreover, though the limit may not be attained in any particular parent ranking, reasons are given in the Appendix for expecting that it cannot be substantially improved upon. The proof is as follows.

Reverting to a finite parent population of N members, we first seek a maximum for Σc_i^2. In terms of the original scores, $c_{ij} = a_{ij}b_{ij}$. Keeping $b_{ij} = \pm 1$, $i \lesseqgtr j$, $b_{ii} = 0$, as before, allow the a_{ij}'s to assume *any* values subject to the conditions

$$\Sigma a_{ij}^2 = N(N-1), \quad \Sigma a_{ij}b_{ij} = c = N(N-1)\tau.$$

The stationary values of Σc_i^2 occur when the a_{ij}'s satisfy the equations

$$b_{ij}(c_i + c_j) - \lambda a_{ij} - \mu b_{ij} = 0,$$

which give, on multiplying by b_{ij} and summing j,

$$c_i = \frac{\mu(N-1)-c}{(N-2-\lambda)}.$$

Thus, unless the c_i's are all to be equal, in which case Σc_i^2 is a minimum, λ and μ must take the values

$$\lambda = N-2, \quad \mu = c/(N-1),$$

and since

$$2\Sigma c_i^2 - \lambda N(N-1) - \mu c = 0,$$

it follows that Σc_i^2 cannot exceed $\frac{1}{2}N(N-1)(N-2)+\frac{1}{2}c^2/(N-1)$. Allowing N to become large, this implies

$$\Sigma r_i^2/N \leqslant \tfrac{1}{2}(1+\tau^2).$$

Hence

$$\operatorname{var} r_i \leqslant \tfrac{1}{2}(1-\tau^2),$$

and so from equation (8·1)

$$\operatorname{var} t \leqslant \frac{2}{n}(1-\tau^2). \tag{9·1}$$

10. Assuming that the sample is large enough for the distribution of t to be normal, the roots τ_1, τ_2 of the equation

$$t-\tau = x\sqrt{\left[\frac{2}{n}(1-\tau^2)\right]}, \tag{10·1}$$

i.e.

$$\tau = \frac{t \pm x\sqrt{\frac{2}{n}}\sqrt{\left(1+\frac{2x^2}{n}-t^2\right)}}{\left(1+\frac{2x^2}{n}\right)}, \tag{10·2}$$

provide confidence limits to τ when t is known, x being the standardized normal deviate corresponding to a given probability of P %. These confidence limits are of course maxima, in the sense that we shall be wrong in *at most* P % of the cases in asserting τ to lie between the calculated limits.

In our proof of the tendency of t to normality it will be necessary to neglect terms of order $n^{-\frac{1}{2}}$, and the sample may have to be rather large for such terms to be small, unless τ itself is small.

The form of equation (9·1) suggests using

$$w = \sin^{-1} t$$

instead of t. To the same order of approximation we can take w as having a normal distribution with mean $\omega = \sin^{-1}\tau$ and standard error not exceeding $\sqrt{(2/n)}$, which is independent of τ. This form is more convenient for assigning confidence limits to t, and for testing the significance of the difference between t_1 and t_2 (whose standard error cannot exceed $\sqrt{[2(1/n_1+1/n_2)]}$), but we have not been able to discover whether the transformation brings the distribution nearer to normality.

11. We now prove that the distribution of t tends for large n to normality whatever the parent ranking, provided that $|\tau|$ is not near unity.

Write $g_{ij} = c_{ij} - c/N^2$ so that $\Sigma g_{ij} = 0$, $g_{ij} = g_{ji}$ and $g_{ii} = -c/N^2 = -(N-1)\tau/N$. The rth moment of $c^{(n)}$ about its mean value is $E[\Sigma^{(n)} g_{ij}]^r$, so consider

$$\sum_n [\Sigma^{(n)} g_{ij}]^r = \sum_n \Sigma^{(n)} g_{ij} g_{kl} g_{uv} \cdots,$$

the summation $\sum_n$ being over all possible sample selections.

The argument used by Daniels (1944) to show that in the null case the distribution of rank correlation in large samples tends to normality can be applied with little modification to the present problem. The proof is therefore sketched here without much detail.

Two essential conditions to be satisfied are that $\Sigma g_{ij} = 0$, which is true by definition, and $\Sigma g_{ij} g_{ik} = O(N^3)$, which is true only if $1 - r^2 = O(1)$, so that the tendency to normality may be expected to break down for high correlations.

The sum $\sum_n$ is evaluated as in § 8 by counting the number of ways in which terms like $g_{ij} g_{kl} g_{uv} \ldots$, and similar terms with tied suffixes, occur. In this way it is expressed as a linear combination of $\Sigma' g_{ij} g_{kl} g_{uv} \ldots$, etc. Every such Σ' is replaceable by the corresponding Σ together with terms containing more tied suffixes which are of lower order in N since they involve fewer summations from 1 to N.

12. First consider the even moments with $r = 2m$. Terms containing more than $3m$ different suffixes must vanish, since in such cases it is impossible to avoid at least one g_{ij} with two free suffixes, and $\Sigma g_{ij} = 0$. For the same reason the only non-vanishing terms with $3m$ different suffixes are those containing expressions like

$$\Sigma g_{ij} g_{ik} g_{lu} g_{lv} g_{pq} g_{pr} \ldots = (\Sigma g_{ij} g_{ik})^m,$$

and terms with fewer different suffixes are of correspondingly lower order in N.

With $3m$ suffixes assigned there are $\binom{N-3m}{n-3m}$ ways of selecting the remaining $n-3m$ members of the sample, and the suffixes can be tied in $\dfrac{(2m)!\,2^{2m}}{m!\,(2!)^m}$ ways to give the same result. Dividing out by $\binom{N}{n}$ and noting that $\binom{N-3m}{n-3m} \Big/ \binom{N}{n} \sim n^{3m}/N^{3m}$ when both N and n are large, the contribution of such terms to μ_{2m}, the $2m$th moment of $c^{(n)}$ about its mean, is found to be

$$\frac{n^{3m}}{N^{3m}} \frac{(2m)!}{m!} 2^m (\Sigma g_{ij} g_{ik})^m,$$

which is of order n^{3m}. Moreover, by the same argument, terms with $f < 3m$ different suffixes add contributions of order n^f which may be neglected.

Hence

$$\mu_{2m} \sim \frac{n^{3m}}{N^{3m}} \frac{(2m)!}{m!} 2^m (\Sigma g_{ij} g_{ik})^m,$$

the neglected terms being relatively $O(n^{-1})$.

13. For the odd moments let $r = 2m+1$. Similar considerations show that the non-vanishing terms of $\sum_n$ cannot have more than $3m+1$ different suffixes, and μ_{2m+1} is therefore of order n^{3m+1}.

Then since $c^{(n)}/n^{\frac{1}{2}}$ has even moments of unit order and odd moments of order $n^{-\frac{1}{2}}$, the odd moments may be neglected to that order. We conclude that $c^{(n)}$ is distributed normally for large n with variance

$$\frac{4n^3}{N^3} \Sigma g_{ij} g_{ik} = 4n^3 \operatorname{var} \tau_i,$$

and t is similarly normal with variance $(4/n) \operatorname{var} \tau_i$.

14. The fact that terms of order $n^{-\frac{1}{2}}$ have to be neglected suggests that the normal approximation only holds good for fairly large samples. This is not surprising since one would expect skewness to be an important property of the distribution of t when τ is not zero, if only for the reason that $|t|$ can never exceed unity. It seems worth while to examine the odd momen in more detail.

The dominant term of the $(2m+1)$th moment has $3m+1$ different suffixes, which can occur as $\Sigma g_{ij}g_{ik}g_{il}(\Sigma g_{uv}g_{uw})^{m-1}$ or $\Sigma g_{ij}g_{ik}g_{jl}(\Sigma g_{uv}g_{uw})^{m-1}$.

Both can be obtained in $$\frac{(2m+1)!\,2^{2(m-1)}\,2^3}{(2!)^{m-1}\,3!(m-1)!} = \frac{(2m+1)!\,2^{m+2}}{3!(m-1)!}$$ distinct ways, and there are $\binom{N-3m-1}{n-3m-1}$ ways of selecting the sample with $3m+1$ suffixes assigned. The $(2m+1)$th moment of $c^{(n)}$ about its mean is therefore

$$\mu_{2m+1} \sim \frac{n^{3m+1}}{N^{3m+1}} \frac{(2m+1)!\,2^{m+2}}{3!(m-1)!} [\Sigma g_{ij}g_{ik}g_{il} + \Sigma g_{ij}g_{ik}g_{jl}] (\Sigma g_{ij}g_{ik})^{m-1},$$

ignoring terms of relative order $O(n^{-1})$. The corresponding moment of t is obtained to the same order on dividing by n^{4m+2}; it depends only on $\operatorname{var} t$ and $\mu_3(t)$, where

$$\mu_3(t) \sim \frac{8}{n^2N^4}[\Sigma g_{ij}g_{ik}g_{il} + \Sigma g_{ij}g_{ik}g_{jl}] = \frac{4}{n^2}\frac{\Sigma g_{ij}(g_i+g_j)^2}{N^4},$$

where $g_i = \sum_{j=1}^{N} g_{ij}$. The distribution of t is thus specified to $O(n^{-1})$ by its first three moments. The moment-generating function of the distribution of t in standard measure is

$$M(z) = \left(1 + \frac{\gamma_1}{3!}z^3\right) e^{-\frac{1}{2}z^2}[1 + O(n^{-1})],$$

where $$\gamma_1 = \mu_3(t)/(\operatorname{var} t)^{\frac{3}{2}} = O(n^{-\frac{1}{2}}),$$

and the frequency distribution of $x = (t-\tau)/\sqrt{(\operatorname{var} t)}$ is*

$$f(x) = \left(1 - \frac{\gamma_1}{3!}\frac{d^3}{dx^3}\right)\frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}}[1 + O(n^{-1})]. \qquad (14{\cdot}1)$$

15. The effect of the γ_1 term in modifying the confidence limits based on normal theory can be seen in the following way. Let ξ be the normal deviate whose chance of being exceeded is $P(\xi)$. The chance of x exceeding ξ is, from (14·1),

$$F(\xi) = P(\xi) + \frac{\gamma_1}{6}(\xi^2 - 1)\frac{e^{-\frac{1}{2}\xi^2}}{\sqrt{(2\pi)}}.$$

If X is the correct limit such that $F(X) = P(\xi)$, it is readily proved by successive approximation that the formula

$$X = \xi + \frac{\gamma_1}{6}(\xi^2 - 1) \qquad (15{\cdot}1)$$

gives the appropriate value of X to $O(n^{-1})$. For example, the 5 and 1 % limits are respectively $\pm 1{\cdot}96 + 0{\cdot}474\gamma_1$ and $\pm 2{\cdot}58 + 0{\cdot}941\gamma_1$.

16. In practice the value of $\operatorname{var} t$ has to be estimated from the sample, and although its standard error can be shown to be $O(n^{-\frac{1}{2}})$ by the kind of argument already used, it is not known how large the sample has to be before the error in estimating the variance can be safely ignored. It is best to use the unbiased formula

$$\operatorname{var} t = \frac{1}{n(n-1)(n-2)(n-3)}\left\{4\Sigma c_i^2 - \frac{2(2n-3)}{n(n-1)}c^2 - 2n(n-1)\right\} \qquad (16{\cdot}1)$$

(which is easily proved) in calculating $\operatorname{var} t$ from the sample, especially if the standard error of the mean value of t from a number of small samples is required.

* Note that the approximation error in $f(x)$ is relatively $O(n^{-1})$, a stronger result than would be obtained from a Gram-Charlier approximation based on the first three moments only.

As the term in γ_1 is a small correction it is perhaps sufficient in moderate samples to take

$$G = \tfrac{1}{2}\Sigma g_{ij}(g_i+g_j)^2 = \tfrac{1}{2}\Sigma c_{ij}(c_i+c_j)^2 - \frac{5c\Sigma c_i^2}{n} + \frac{3c^2}{n^3}, \tag{16·2}$$

and

$$\mu_3(t) = \frac{8}{n^6}G, \quad \gamma_1 = \mu_3(t)/(\text{var}\, t)^{\frac{3}{2}}, \tag{16·3}$$

where the first term in G is the sum of $c_{ij}(c_i+c_j)^2$ over all values of $i>j$. The unbiased formula for $\mu_3(t)$ involves some rather tedious computation.

17. To illustrate the methods of the paper we consider an actual example.

A set of thirty wool samples were visually graded in order of fibre fineness by three assessors. The mean fibre diameter for each wool sample was also determined by direct measurement. Table 1 shows the measured order (M) compared with that of the three assessors (A, B, C), in ascending order of experience.

Table 1

M	A	B	C	M	A	B	C
1	5	2	1	16	12	14	16
2	4	5	2	17	10	18	15
3	9	6	6	18	30	21	25
4	3	1	3	19	22	26	24
5	6	7	4	20	16	22	19
6	2	4	5	21	21	16	18
7	15	19	10	22	29	20	23
8	18	3	12	23	28	25	22
9	8	8	7	24	19	27	26
10	11	9	8	25	23	28	21
11	17	13	9	26	20	23	27
12	13	10	11	27	7	24	20
13	24	17	17	28	26	29	28
14	14	12	14	29	27	15	30
15	1	11	13	30	25	30	29

The method of working will be seen from the c_{ij} matrix for the MA correlation shown in Table 2.

The correlations of the assessors' orders with the measured order are found to be

$$t_A = 0{\cdot}490, \quad t_B = 0{\cdot}724, \quad t_C = 0{\cdot}816.$$

(i) Consider first the maximum confidence limits given by (10·2). The 5 % limits are

$$-0{\cdot}02 < t_A < 0{\cdot}80, \quad 0{\cdot}23 < t_B < 0{\cdot}92, \quad 0{\cdot}34 < t_C < 0{\cdot}96.$$

Again, using the transformation $w = \sin^{-1}t$, the 5 % limits are

$$0{\cdot}01 < t_A < 0{\cdot}85, \quad 0{\cdot}30 < t_B < 0{\cdot}97, \quad 0{\cdot}45 < t_C < 0{\cdot}99.$$

The values of w are $w_A = 0{\cdot}512, \quad w_B = 0{\cdot}810, \quad w_C = 0{\cdot}954.$

The greatest difference is 0·442, and the upper limit to its standard error is $\sqrt{(4/n)} = 0{\cdot}365$, so on these grounds the difference between A and C would not be judged significant.

The 5 % limits are very wide, and the lack of significance is disappointing since C was known to be an expert appraiser while A is relatively inexperienced, and one would have expected an obvious difference between them.

(ii) The variances estimated from the unbiased formula (16·1) are

$$\operatorname{var} t_A = 0{\cdot}006630, \quad \operatorname{var} t_B = 0{\cdot}005067, \quad \operatorname{var} t_C = 0{\cdot}002198.$$

The estimated standard errors are therefore

$$s_A = 0{\cdot}081, \quad s_B = 0{\cdot}071, \quad s_C = 0{\cdot}047.$$

The 5 % confidence limits, assuming normality, are

$$0{\cdot}33 < t_A < 0{\cdot}65, \quad 0{\cdot}58 < t_B < 0{\cdot}86, \quad 0{\cdot}72 < t_C < 0{\cdot}91.$$

Table 2

c_{ij}																												c_i
0	−	+	−	+	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
−	0	+	−	+	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
+	+	0	−	−	−	+	+	−	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	−	+	+	17
−	−	−	0	+	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	19
+	+	−	+	0	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23
−	−	−	−	−	0	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	17
+	+	+	+	+	0	+	−	−	+	−	+	−	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	13
+	+	+	+	+	+	+	0	−	−	−	−	+	−	−	−	−	+	+	−	+	+	+	+	+	−	+	+	9
+	+	−	+	+	+	−	−	0	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	−	+	+	19
+	+	+	+	+	+	−	−	+	0	+	+	+	+	−	+	−	+	+	+	+	+	+	+	+	−	+	+	19
+	+	+	+	+	+	+	−	+	+	0	−	+	−	−	−	−	+	+	−	+	+	+	+	+	−	+	+	13
+	+	+	+	+	+	+	−	−	+	+	0	+	+	−	−	−	+	+	+	+	+	+	+	+	−	+	+	15
+	+	+	+	+	+	+	+	+	+	+	+	0	−	−	−	−	+	−	−	−	+	+	−	−	−	−	+	+	7
+	+	+	+	+	+	−	−	+	+	−	+	−	0	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	13
−	−	−	−	−	−	−	−	−	−	−	−	−	−	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1
+	+	+	+	+	+	−	−	+	+	−	−	−	−	+	0	−	+	+	+	+	+	+	+	+	+	−	+	+	13
+	+	+	+	+	+	+	−	−	+	−	−	−	−	+	−	0	+	+	+	+	+	+	+	+	+	−	+	+	11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	−	−	−	−	−	−	−	−	−	−	−	5
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	0	−	−	+	+	−	+	−	−	+	+	15
+	+	+	+	+	+	+	−	+	+	−	+	−	+	+	+	+	−	−	0	+	+	+	+	+	+	−	+	+	17
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	0	+	+	−	+	+	−	+	+	19
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	0	−	−	−	−	−	−	−	11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	0	−	−	−	−	−	−	11
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	−	−	−	0	+	+	−	+	+	15
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	−	−	+	0	−	−	+	+	17
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	−	−	+	−	0	−	+	+	15
+	+	−	+	+	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	0	+	+	−11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	0	−	21
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	+	0	21
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	−	−	19

$c = 426$

$n = 30$

$\Sigma c_i^2 = 7470$

Moreover, we should judge A and C to be significantly different at the 1 % level, and A and B at the 5 % level. How far these conclusions are valid depends, of course, on the accuracy of the variance estimates, but the conclusions seem to agree with what might have been expected from prior knowledge of the assessors' capabilities.

(iii) The values of γ_1 calculated from (16·2) and (16·3) are

$$\gamma_1(A) = -0{\cdot}32, \quad \gamma_1(B) = -0{\cdot}35, \quad \gamma_1(C) = -0{\cdot}38.$$

The distributions would not appear to be very skew, and the distribution of the difference of two t's is probably nearly normal. The adjusted 5 % limits are, from (15·1),

$$0{\cdot}32 < t_A < 0{\cdot}64, \quad 0{\cdot}57 < t_B < 0{\cdot}85, \quad 0{\cdot}72 < t_C < 0{\cdot}90.$$

APPENDIX

1. The question arises whether a particular parental form exists for which the variance of t assumes the upper limit $2(1-r^2)/n$. We surmise, though we cannot prove, that the maximum possible variance is attained when the parent ranking has a 'canonical' form obtained in the following way. Consider again the ranking

5 2 3 1 6 7 8 9 4.

The number of positive pairs S is 26, so that $t = 0{\cdot}44$. Let us transform this so as to bring the 1 to the beginning of the ranking but move the 9 so as to preserve the number S at 26. The 1 passes over three members to go to the beginning and hence adds 3 to the score. The 9 must, therefore, proceed to the left over three numbers so as to subtract 3 from the score and we reach

1 5 2 3 9 6 7 8 4.

Now operate similarly with 2 and 9, reaching

1 2 5 9 3 6 7 8 4.

Had our 9 been contiguous to the 1 and incapable of moving farther to the left we should have moved the 8 and so on. Proceeding with the process by moving back the 3 and the 9 and 8 we reach

1 2 3 9 5 6 8 7 4,

and again

1 2 3 4 9 8 7 6 5.

All the lower numbers 1 to 4 are in the right order and the remainder are in the inverse order. We call this ranking the 'canonical' order for given S (or t). It is not always possible to reduce a given ranking to canonical order, but there cannot be more than one individual out of place.

2. Consider the effect of a series of transformations leading to the canonical form. The first process, that of moving 1 and 9, will increase the value of S for some samples involving 1 but not 9 (leaving the others unchanged), will decrease the value of S for some samples involving 9 but not 1 (leaving the others unchanged), and will, in general, not alter those involving both 1 and 9. Similarly for 2 and 8, and so on. The effect of the transformation is thus to increase the values of S containing the lower numbers 1, 2, 3, etc., and to decrease those containing 9, 8, 7, etc. These values of S are themselves, in the canonical form, the greatest or least as the case may be. Consequently the progress to the canonical form is accompanied by increases in the number of high values of S and increases in the number of lower values, and one might expect the spread of the distribution to tend to a maximum. In the example quoted, the distributions of S in samples of 3 for the successive rankings are:

Values of S	Frequencies f				
0	2	3	3	6	10
1	15	13	16	10	—
2	34	35	29	32	40
3	33	33	36	36	34
Totals	84	84	84	84	84

The sums ΣfS are all equal to 182. The sums ΣfS^2 are respectively 448, 450, 456, 462 and 466, showing the canonical ranking to have the largest variance of the five.

3. There is, however, another way of carrying out this process. If the parent ranking is inverted, τ becomes $-\tau$, but the variance of samples of n drawn from the inverted ranking remains the same, by symmetry. We may then reduce the inverted ranking to its canonical form and reinvert it so that its coefficient is again τ. This ranking we call the inverse canonical form. It will be shown that for large N, when $\tau > 0$ the inverse canonical form yields a larger variance for t than the direct canonical form.

Even in the example already quoted, the inverse canonical ranking (with one member out of place) is 3 4 2 5 6 7 8 9 1,
which has a distribution

Values of S	f
0	2
1	27
2	10
3	45
Total	84

The sum ΣfS^2 is now 472, which is greater than the previous maximum 466.

4. Consider the canonical case when there are N members altogether, R at the beginning in the right order, and $N-R$ in the inverse order. If we select $n-j$ members from the R and j from the $N-R$ the value of S for the sample of n is $\frac{1}{2}n(n-1)-\frac{1}{2}j(j-1)$, and the relative frequency of $U = \frac{1}{2}n(n-1)-S$ is $\binom{R}{n-j}\binom{N-R}{j}\Big/\binom{N}{n}$. Now suppose that N tends to infinity and R/N to the ratio p. The relative frequency of $U = \frac{1}{2}j(j-1)$ tends in the limit to

$$\binom{n}{j}p^{n-j}q^j,$$

where $q = 1-p$. The mean value of U is then

$$\sum_0^n \tfrac{1}{2}j(j-1)\binom{n}{j}p^{n-j}q^j = \tfrac{1}{2}n(n-1)q^2,$$

and since

$$t = 1-\frac{2U}{\frac{1}{2}n(n-1)},$$

we must have

$$q = \{\tfrac{1}{2}(1-\tau)\}^{\frac{1}{2}}. \qquad (4{\cdot}1\,\mathrm{A})$$

The variance of U is

$$\operatorname{var} U = n(n-1)\,pq^2\{nq-\tfrac{1}{2}(1-3q)\}, \qquad (4{\cdot}2\,\mathrm{A})$$

and so

$$\operatorname{var} t = 16pq^2\{nq-\tfrac{1}{2}(1-3q)\}/n(n-1). \qquad (4{\cdot}3\,\mathrm{A})$$

5. If now the inverted parent ranking is reduced to canonical form, giving ratios p', q' corresponding to p and q, we shall have

$$q' = \sqrt{[\tfrac{1}{2}(1+\tau)]} \qquad (5{\cdot}1\,\mathrm{A})$$

and

$$\operatorname{var} t' = 16p'q'^2\{nq'-\tfrac{1}{2}(1-3q')\}/n(n-1). \qquad (5{\cdot}2\,\mathrm{A})$$

Then since $q^2+q'^2 = 1$,

$$\operatorname{var} t' - \operatorname{var} t = \frac{16(n+2)}{n(n-1)}(q'-q)(1-q)(1-q'). \qquad (5{\cdot}3\,\mathrm{A})$$

When τ is positive, $q' > q$ and $\operatorname{var} t'$ exceeds $\operatorname{var} t$.

This result suggests that the maximum variance may be attained by the inverse canonical ranking when $\tau > 0$ and by the direct canonical ranking when $\tau < 0$. With this choice of parent ranking the variance of t for large n is

$$\operatorname{var} t \sim \frac{4\sqrt{2}}{n}(1+|\tau|)^{\frac{1}{2}}\{1-\sqrt{[\tfrac{1}{2}(1+|\tau|)]}\}. \qquad (5{\cdot}4\,\mathrm{A})$$

It is interesting to compare (5·4 A) with our upper limit of $2(1-\tau^2)/n$. Their ratio is $\{2(1+|\tau|)\}^{\frac{1}{2}}/[1+\{\frac{1}{2}(1+|\tau|)\}^{\frac{1}{2}}]$, which varies from $2(\sqrt{2}-1) = 0{\cdot}83$ when $\tau = 0$ to 1 when $\tau = 1$. Evidently the upper limit to the variance cannot be much improved, since an actual ranking has been found whose variance approximates to it for all values of τ, when n is not too small.

REFERENCES

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. 2nd ed. 1945. London: Charles Griffin and Co.

TESTING FOR NORMALITY

By R. C. GEARY, *Cambridge University Department of Applied Economics*

1. Introduction

The present communication, one of a series, has two main objectives:

(1) To show that probabilities derived from the well-known analyses of variance and other 'small sample' tables, which postulate universal normality, may differ seriously from the true probabilities when the universes are non-normal, even, in some cases, when the degree of non-normality is not considerable.

(2) To determine the most efficient tests of normality from a wide field of alternative symmetrical tests.

It may be useful to summarize very briefly previous work in so far as it is strictly relevant to this study.* The modern theory may be regarded as having been initiated by Karl Pearson who, in 1895, found the first approximation (i.e. to n^{-1}) to the variances and covariance of $\sqrt{b_1}$ and b_2 for samples drawn at random from any universe and, assuming that the $\sqrt{b_1}$ and b_2 were distributed jointly with normal probability, constructed 'probability ellipses' from which the probability of the same values occurring, had the universe, in fact, been normal, could be inferred very approximately. A considerable advance in moment determination was made by C. C. Craig (1928). In 1929, R. A. Fisher, in inventing cumulants, simple functions of the sample moments, and formulating rules for finding their semi-invariants, developed incidentally a technique for expanding to several terms in $1/n$ the moments of $\sqrt{b_1}$ and b_2 when the universe was normal. This paper was followed soon after by another (1930), fundamental for all succeeding work on this subject, in which R. A. Fisher ingeniously applied combinatorial technique to the finding of exact values of the moments of normal $\sqrt{b_1}$ and b_2, and gave *inter alia* the values of the second, fourth and sixth moments of $\sqrt{b_1}$ and of the first three moments of b_2. The fourth semi-invariant, together with many other normal semi-invariants of b_2, was determined by J. Wishart in 1930, and a further advance in R. A. Fisher's technique was made jointly by R. A. Fisher & J. Wishart in 1930. In 1932 Joseph Pepper gave the eighth normal moment of $\sqrt{b_1}$. Using R. A. Fisher's rules C. T. Hsu and D. N. Lawley in 1940 gave the exact values for normal random samples of the fifth and sixth moments of b_2. Using a method due to R. C. Geary (1933) (applying C. C. Craig's ideas (1928) to the normal problem), R. C. Geary & J. P. G. Worlledge have recently (1946) found the seventh moment of b_2.

So much for moment determination. In 1930, E. S. Pearson used appropriate Pearson-type curves, applied to R. A. Fisher's (1929) approximations of the semi-invariants, to find approximate frequency distributions of $\sqrt{b_1}$ and b_2. From the frequency distributions he computed a table of 1 % and 5 % probability points at intervals for n from 50 to 5000 for $\sqrt{b_1}$ and for n from 100 to 5000 for b_2.

Since at the time the prospect seemed remote of determining the frequency of normal b_2 on which reliance could be reposed for samples of moderate sizes, R. C. Geary (1935)† suggested that the ratio, a, of mean deviation to standard deviation computed from the origin

* An excellent account of the development of moment theory up to the year 1930 was given by J. Wishart (1930).

† The author was informed by M. Fréchet that this test was suggested by Bertrand, but has been unable to check the reference.

might be used as a test of normality, and gave the 1 and 5 % probability points for this test at intervals for normal samples of 6–100. E. S. Pearson compared experimentally Geary's test with b_2 and suggested, for samples so large that comparison could safely be made, that b_2 was probably somewhat more sensitive than a, a suggestion which will be examined theoretically in this communication. In 1935 also, R. C. Geary showed that there was a high (negative) correlation for normal samples between $a(1)$ (see 3·1) and b_2 for normal samples, and argued therefrom that the former should be nearly as efficient as b_2. In 1936, R. C. Geary gave a table of 1, 5 and 10 % probability points of $a(1)$ at intervals for samples of 11–1001. In 1938, a brochure by R. C. Geary & E. S. Pearson was published by the Biometrika Office entitled *Tests of Normality*, giving tables and diagrams of probability points of $a(1)\sqrt{b_1}$ and b_2. There is considerable literature dealing with the effect of universal non-normality on the normal tests, mostly by way of particular numerical examples: a selection of papers on this subject is included in the list of references at the end of the paper.

2. Effect of non-normality

(a) *The z-test*

The effect of universal non-normality will first be considered in relation to the z-test. If $x_1, x_2, \ldots, x_{n'}$ and $y_1, y_2, \ldots, y_{n''}$ are two independent samples drawn at random from the same universe (normal or non-normal) it is easy to show that, if

$$z = \tfrac{1}{2}\log\frac{n''-1}{n'-1}\frac{\sum_{i=1}^{n'}(x_i-\bar{x})^2}{\sum_{i=1}^{n''}(y_i-\bar{y})^2} = \tfrac{1}{2}\log\frac{s'^2}{s''^2}, \tag{2·1}$$

then

$$\sigma_z^2 = \frac{(\beta_2-1)}{4}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2, \tag{2·2}$$

when both n' and n'' are so large that terms in n' and n'' of degree less than -1 are regarded as negligible. This is an obvious generalization of the approximate formula given by R. A. Fisher* for normal samples, namely,

$$\sigma_z^2 = \frac{1}{2}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2^0. \tag{2·3}$$

It may be useful also to give formulae for the first and second moments from zero for z when the two random samples are drawn not necessarily from the same universes, though both universes have mean zero and the same variance λ_2:

$$\left.\begin{aligned}
2M_1' &= -\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4'}{n'}+\frac{2\lambda_2^2}{n'-1}\right)+\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4''}{n''}+\frac{2\lambda_2^2}{n''-1}\right)+\frac{1}{3\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}-\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}-\frac{\lambda_4''}{n''^2}\right)\right.\\
&\qquad\left.+4\left(\frac{\lambda_3'^2}{n'^2}-\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}-\frac{1}{n''^2}\right)\right]-\frac{3}{4\lambda_2^4}\left[\frac{(\lambda_4'+2\lambda_2^2)^2}{n'^2}-\frac{(\lambda_4''+2\lambda_2^2)^2}{n''^2}\right]+\ldots,\\
4M_2' &= \frac{1}{\lambda_2^2}\left[\left(\frac{\lambda_4'}{n'}+\frac{\lambda_4''}{n''}\right)+2\lambda_2^2\left(\frac{1}{n'-1}+\frac{1}{n''-1}\right)\right]\\
&\quad-\frac{1}{\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}+\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}+\frac{\lambda_4''}{n''^2}\right)+4\left(\frac{\lambda_3'^2}{n'^2}+\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}+\frac{1}{n''^2}\right)\right]\\
&\quad+\frac{1}{12\lambda_2^4}\left[\frac{33}{n'^2}(\lambda_4'+2\lambda_2^2)^2+\frac{33}{n''^2}(\lambda_4''+2\lambda_2^2)^2-\frac{6}{n'n''}(\lambda_4'+2\lambda_2^2)(\lambda_4''+2\lambda_2^2)\right]+\ldots
\end{aligned}\right\} \tag{2·4}$$

* *Statistical Methods for Research Workers*, 8th ed. p. 219.

where the λ's indicate semi-invariants of the two universes of the orders indicated. In these formulae, in effect, terms to order -2 in n', n'' are retained.

When both samples are large the frequency distribution of z will approach normality provided that μ_4 is finite. The effect of universal kurtosis can accordingly be assessed in a very rudimentary manner from (2·2) and (2·3). The z-deviate ζ corresponding to, say, the $2\frac{1}{2}$ % normal probability point is

$$\zeta = 1{\cdot}9600\sqrt{M_2^0}. \tag{2·5}$$

If, however, the universe were not normal and had, *in fact*, a variance M_2 with $\beta_2 \neq 3$, the *actual* probability of a deviation in excess of ζ in absolute value would be, not 0·05, but the normal probability appropriate to a unit variance deviate of $\zeta M_2^{-\frac{1}{2}}$. On this consideration the actual probabilities for different values of β_2, where the assumed probability is 0·05, are shown in the fifth column of Table 1.

Table 1. *Effect on probability of z of change in universal kurtosis, for large samples*

β_2	M_2^0/M_2	$\sqrt{(M_2^0/M_2)}$	1·9600 $\sqrt{(M_2^0/M_2)}$	Actual probability
1·5	4	2	3·9200	0·000089
2	2	1·4142	2·7718	0·0056
2·5	1·3333	1·1547	2·2632	0·024
3	**1**	**1**	**1·9600**	**0·050**
3·5	0·8000	0·8944	1·7530	0·080
4	0·6667	0·8165	1·6003	0·110
4·5	0·5714	0·7559	1·4816	0·138
5	0·5000	0·7071	1·3859	0·166
5·5	0·4444	0·6667	1·3065	0·191
6	0·4000	0·6325	1·2397	0·215

The table shows that, if the universe from which the samples are drawn has $\beta_2 = 6$, the true probability is about 1 in 5 instead of the assumed 1 in 20. It is, of course, true that universes with so large a kurtosis are unusual. This view cannot be held of the range 2·5–4 for β_2 in which the probability, assumed to be 0·05, can be anything, in fact, from 0·024 to 0·110. Accordingly, if universal kurtosis is markedly negative, use of the standard table masks significant differences; if kurtosis is positive the standard table exaggerates these differences. Unless systematic tests have established that kurtosis is negligible the standard table should not be used for testing significant differences in variance.

The foregoing analysis gives a theoretical explanation of the striking experimental results of E. S. Pearson (1931*b*) working, however, with a test function

$$x = \sum_{i=1}^{n'} (x_i - \bar{x})^2 \Big/ \left\{ \sum_{i=1}^{n'} (x_i - \bar{x})^2 + \sum_{i=1}^{n''} (y_i - \bar{y})^2 \right\}$$

and with sample sizes $n' = 5$ and $n'' = 20$, smaller than those contemplated in the present analysis. With 500 samples Pearson showed that when the frequency at the two tails together expected from normal theory was 15·4 (= probability 0·0308) the frequencies actually found in symmetrical universes with $\beta_2 = 2{\cdot}5$, 4·1 and 7·1 respectively were 7, 39 and 47, equivalent to probabilities of 0·014, 0·078 and 0·094.

If tests of normality indicate universal kurtosis, either of two courses might be adopted:

(i) Assume that z is normally distributed with variance M_2 computed from (2·2) with (β_2-3) estimated as k_4/k_2^2 from the sample, k_2 and k_4 being R. A. Fisher's (1929) cumulant functions.

(ii) Enter the standard table, not with z computed from the samples but with $z\sqrt{(M_2^0/M_2)}$, estimating M_2 as in (i).

Both of these procedures are, of course, open to the objection that, unless the samples are extremely large the estimate of β_2 is unlikely to be accurate; the real β_2 might be larger or smaller than the estimate. Any probabilistic inferences should accordingly be accepted with reserve.

It is fortunate that the condition specified in the foregoing paragraphs, namely, that the numbers in the two samples are both large, rarely applies in practical applications. It more usually happens that the number of classes is small, whereas the number per class is relatively large. In this case E. S. Pearson (1931*b*) has shown the first approximation to σ_z^2 is independent of β_2, from which he inferred that the actual probability when the total number of samples was large was inconsiderably influenced by kurtosis. In view of the foregoing analysis it seemed to the writer desirable to carry the inquiry a stage further.

Suppose, then, that k samples are drawn at random from the same universe, n_j in the jth sample, the total $\sum_j n_j = n$. It is assumed that n is so large that terms in n^{-2} are negligible, that the number of samples k is small, and that all the n_j are of the same order of magnitude as n, i.e. that if

$$n_j = \pi_j n, \quad \sum_{j=1}^{k} \pi_j = 1, \tag{2·6}$$

none of the π_j is negligibly small.

Using R. A. Fisher's cumulant notation with subscript to indicate the sample from which the cumulants were computed, the mean for the jth sample is written k_{1j} and its variance k_{2j}. Then

$$z = \tfrac{1}{2}\log\frac{X}{Y}, \tag{2·7}$$

where
$$(k-1)\,X = \sum_j n_j(k_{1j}-k_1)^2 = \sum n_j k_{1j}^2 - \frac{1}{n}\Sigma^2 n_j k_{1j},$$

so that
$$\frac{k-1}{n}X = \sum \pi_j(1-\pi_j)\,k_{1j}^2 - 2\sum_{j<j'}\sum \pi_j\pi_{j'}k_{1j}k_{1j'},$$

and
$$(n-k)\,Y = \sum_j (n_j-1)\,k_{2j},$$

so that
$$Y = \Sigma\phi_j k_{2j},$$

where
$$\phi_j = \frac{n_j-1}{n-k}.$$

Without loss of generality let the universal mean be zero and the variance unity. It may easily be shown that

$$EX = EY = 1.$$

Set
$$w = \frac{X}{Y} = \frac{\bar{X}+(X-\bar{X})}{\bar{Y}+(Y-\bar{Y})} = \{1+(X-1)\}\{1+(Y-1)\}^{-1}.$$

Then
$$\left.\begin{aligned} w &= \{1+(X-1)\}\{1-(Y-1)+(Y-1)^2-(Y-1)^3+\ldots\},\\ w^2 &= \{1+(X-1)\}^2\{1-2(Y-1)+3(Y-1)^2-4(Y-1)^3+\ldots\}.\end{aligned}\right\} \tag{2·8}$$

We shall compute the approximate values of Ew and Ew^2, i.e. the values to order n^{-1}; the symbol $\doteq$ denotes 'equal to, to approximation required'. From values of the variances and covariances given by E. S. Pearson (1931*b*) in his equations (9)–(11), we have

$$\left.\begin{aligned} E(X-1)^2 &= \frac{2}{k-1}+(1-2k+\alpha_{-1}-1)\frac{\lambda_4}{n(k-1)^2},\\ E(X-1)(Y-1) &\doteq \frac{\lambda_3}{n},\\ E(Y-1)^2 &\doteq \frac{\lambda_4+2}{n}, \end{aligned}\right\} \tag{2·9}$$

with $$\alpha_c = \sum_j \pi_j^c.$$

We require $$\left(\frac{k-1}{n}\right)^2 X^2 \doteq \sum_j \pi_j^3(1-\pi_j)^2 k_{1j}^4 - 4\sum_j\sum_{j'} \pi_j^3(1-\pi_j)\pi_{j'} k_{1j}^3 k_{1j'}$$
$$+2\sum_j\sum_{j'}\pi_j\pi_{j'}(1-\pi_j-\pi_{j'}+3\pi_j\pi_{j'})k_{1j}^2k_{2j'}^2 - 4\sum_j\sum_{j'}\sum_{j''}\pi_j\pi_{j'}\pi_{j''}(1-3\pi_j)k_{1j}^2k_{1j'}k_{1j''},$$

$$Y-1 = \Sigma\phi_j(k_{2j}-1) = \Sigma\phi_j k'_{2j}, \quad \text{say,}$$

remembering that, by definition of cumulants,

$$Ek_{2j} = \lambda_2 = 1.$$

Also $$(Y-1)^2 = \Sigma\phi_j^2 k_{2j}'^2 + 2\sum_{j>j'}\sum\phi_j\phi_{j'}k'_{2j}k'_{2j'}.$$

It will be useful for what follows to note that

$$\phi_j \doteq \pi_j.$$

Using R. A. Fisher's formulae (1929) for formation of joint semi-invariants of k_1 and k_2, and noting that the k samples are independent, we find from the foregoing

$$\left.\begin{aligned} n(k-1)\,EX(Y-1)^2 &\doteq (k-1)(\lambda_4+2),\\ n(k-1)^2\,EX^2(Y-1) &\doteq 2(k^2-1)\lambda_4,\\ n(k-1)^2\,EX^2(Y-1)^2 &\doteq (k^2-1)(\lambda_4+2). \end{aligned}\right\} \tag{2·10}$$

Then, from (2·8), (2·9), (2·10),

$$\left.\begin{aligned} Ew &\doteq 1+\frac{2}{n},\\ Ew^2 &\doteq \frac{k+1}{k-1}+\frac{1}{n(k-1)^2}\{6(k^2-1)-(k^2+2k-2-\alpha_{-1})\lambda_4\}. \end{aligned}\right\} \tag{2·11}$$

These are the formulae required. It will be noted

(i) that the terms free of n^{-1} are independent of λ_4, which is equivalent to E. S. Pearson's result (1931*b*);

(ii) that the formulae (2·11) agree with the normal values

$$\left.\begin{aligned} E_0 w &= \left(1-\frac{2}{n-k}\right)^{-1} \doteq 1+\frac{2}{n},\\ E_0 w^2 &= \frac{k+1}{k-1}\left(1-\frac{2}{n-k}\right)^{-1}\left(1-\frac{4}{n-k}\right)^{-1} \doteq \frac{k+1}{k-1}\left(1+\frac{6}{n}\right), \end{aligned}\right\} \tag{2·12}$$

to n^{-1} when $\lambda_4 = 0$;

(iii) the approximations at (2·11) are free of λ_3.

The approximations at (2·11) tend to confirm E. S. Pearson's result that, when n is large compared with k, the effect of universal kurtosis is unimportant. It would be useful, however, to compute the approximate true probability for different values of k, n, λ_4 and α_{-1}. For this and for subsequent work the following lemma* will be found useful:

If $f(x)$ and $\phi(x)$ are two frequency densities with semi-invariants L_m and L'_m $(m = 1, 2, \ldots)$, respectively, then, formally,

$$f(x) = \exp\left\{\sum_{m=1}^{\infty} \frac{(L_m - L'_m)}{m!}\left(-\frac{d}{dx}\right)^m\right\}\phi(x). \tag{2·13}$$

For the present application take as generating function ϕ the frequency distribution of w in the normal case, i.e.

$$\phi(w) = \frac{\left(\frac{n-3}{2}\right)!\left(\frac{k-1}{n-k}\right)^{\frac{1}{2}(k-1)}}{\left(\frac{k-3}{2}\right)!\left(\frac{n-k-2}{2}\right)!}\, w^{\frac{1}{2}(k-3)}\left\{1+\frac{(k-1)w}{(n-k)}\right\}^{-\frac{1}{2}(n-1)}, \tag{2·14}$$

and, from (2·11),

$$L_2 - L'_2 \doteq -\frac{(k^2+2k-2-\alpha_{-1})\lambda_4}{n(k-1)^2}. \tag{2·15}$$

Assume that

$$L_m - L'_m \doteq 0 \quad (m \neq 2).$$

Then if the 'normal theory' probability corresponding to the sample value w be p, the approximate 'true' probability, subject to (2·15), will be about $(p+p')$, where p' is given by

$$p' = \frac{(L_2 - L'_2)}{2}\int_w^{\infty} \phi''(w)\,dw = -\frac{(L_2 - L'_2)}{2}\phi'(w). \tag{2·16}$$

The term p', of course, merely corrects for the non-normal term in n^{-1} in the variance of z; it takes no account of corrections due to terms of higher (negative) orders in n or even of non-normal terms in n^{-1} in semi-invariants L_m $(m > 2)$. The calculation is designed merely to show whether the standard table probability requires correction for universal kurtosis; this will appear if p' is of the order of magnitude of p.

(*b*) *The t-test*

In Geary's 1936 paper the expansion to terms in n^{-2} of the first *four* moments of t, where

$$t = n^{\frac{1}{2}}k_1/k_2^{\frac{1}{2}}, \tag{2·17}$$

were given. Following are the first *six* semi-invariants L of t to the same approximation as in the earlier paper:

$$\left.\begin{aligned}
L_1 &\doteq -\frac{1}{n^{\frac{1}{2}}}\left\{\frac{\lambda_3}{2}+\frac{3}{16n}(2\lambda_3 - 2\lambda_5 + 5\lambda_3\lambda_4)\right\}+\ldots,\\
L_2 &\doteq 1+\tfrac{1}{4}(8+7\lambda_3^2)\,n^{-1}+(6-2\lambda_4-\tfrac{3}{8}\lambda_3^2-\tfrac{45}{8}\lambda_3\lambda_5+\tfrac{177}{16}\lambda_3^2\lambda_4)\,n^{-2},\\
L_3 &\doteq -2\lambda_3 n^{-\frac{1}{2}}-(9\lambda_3-3\lambda_5+\tfrac{15}{4}\lambda_3\lambda_4+\tfrac{93}{8}\lambda_3^3)\,n^{-\frac{3}{2}},\\
\dagger L_4 &\doteq (6-2\lambda_4+12\lambda_3^2)\,n^{-1}+(54-18\lambda_4+4\lambda_6+75\lambda_3^2-63\lambda_3\lambda_5-6\lambda_4^2+81\lambda_3^2\lambda_4+\tfrac{999}{8}\lambda_3^4)\,n^{-2},\\
L_5 &\doteq -(60\lambda_3-6\lambda_5-20\lambda_3\lambda_4+105\lambda_3^3)\,n^{-\frac{3}{2}},\\
L_6 &\doteq (240-120\lambda_4+577\tfrac{1}{2}\lambda_3^2+16\lambda_6-210\lambda_3\lambda_5-150\lambda_3^2\lambda_4+1200\lambda_3^4)\,n^{-2}.
\end{aligned}\right\} \tag{2·18}$$

Throughout this subsection we take

$$\lambda_m = \lambda'_m/\lambda_2'^{\frac{1}{2}m},$$

* Due to Charlier and termed the "Differential Series" by the Scandinavian School.

† 1936 formula corrected.

where the λ'_m are the semi-invariants of the parent universe. For these expressions terms in $n^{-\frac{1}{2}}$ are neglected. They were derived from the moments (from zero) M'_i of t, which were obtained by the method described in the 1936 paper. It will be noted that, to the approximation used, the expressions involve only the first six semi-invariants of the parent universe. When the parent universe is normal all the λ_i $(i > 2)$ are zero. The magnitude of the numerical coefficients in the foregoing approximate expressions for the L_i indicate that, when the universal values of the λ_i, particularly those of uneven order, are not very small, the frequency distribution of t may differ appreciably from the classical Gosset-Fisher (1908, 1925) distribution.

The formal Gram-Charlier expression for the frequency of t could, of course, be written down at once from (2·18). It is doubtful, however, if the Gaussian can be regarded as the most appropriate generating function for the frequency of t because, even when the parent universe is normal, the semi-invariants T'_{2m} of the higher even orders are large for moderate values of n. For example,

$$L'_4/L'^2_2 = 6/(n-5); \quad L'_6/L'^3_2 = 240/(n-5)(n-7).$$

It is proposed to use (2·13) for finding the approximate frequency with

$$\phi(t) = T(t; n) = \left(\frac{n-2}{2}\right)! \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \Big/ \left(\frac{n-3}{2}\right)! \, (\pi\overline{n-1})^{\frac{1}{2}}, \tag{2·19}$$

the Gosset-Fisher frequency. Let

$$T'_1(t; n) = \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \tag{2·20}$$

It can easily be shown that the rth derivative (in t) of T'_1 is

$$T'^{(r)}_1(t; n) = (-)^r \frac{(n+r-1)!}{(n-1)!\,(n-1)^r} \left\{ t^r - n_1 \frac{r(r-1)}{2} t^{r-2} + n_2 \frac{r(r-1)(r-2)(r-3)}{2.4} t^{r-4} \right.$$
$$\left. - n_3 \frac{r(r-1)(r-2)(r-3)(r-4)(r-5)}{2.4.6} t^{r-6} + \ldots \right\} \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}(n+2r)}, \tag{2·21}$$

with $$n_1 = \frac{n-1}{n+1}, \quad n_2 = \frac{(n-1)^2}{(n+1)(n+3)}, \quad n_3 = \frac{(n-1)^3}{(n+1)(n+3)(n+5)}, \quad \text{etc.}$$

Note that (2·21) assumes the Hermite form when $n = \infty$.

The theory will now be applied to particular examples using in all cases $n = 10$. The universes will be assumed to belong to the Karl Pearson system, so that (M. G. Kendall, 1941) the values of λ_5 and λ_6 can be derived (given λ_3 and λ_4) from the following equations:

$$\left.\begin{aligned} (1+4\eta)\lambda_3 + 2\xi &= 0, \\ (1+5\eta)\lambda_4 + 3\xi\lambda_3 + 6\eta &= 0, \\ (1+6\eta)\lambda_5 + 4\xi\lambda_4 + 24\eta\lambda_3 &= 0, \\ (1+7\eta)\lambda_6 + 5\xi\lambda_5 + 10\eta(4\lambda_4 + 3\lambda_3^2) &= 0. \end{aligned}\right\} \tag{2·22}$$

From the first two equations

$$\eta = (2\lambda_4 - 3\lambda_3^2)/(-10\lambda_4 + 12\lambda_3^2 - 12),$$

which, substituted in the first equation of (2·22), gives ξ. The values of ξ and η, substituted in the third and fourth equations, give λ_5 and λ_6. From (2·18), the L'_i being the semi-invariants

when the parent universe is normal (i.e. the values found when all the λ's are set equal to zero),

$$\left.\begin{aligned} L_1 - L_1' &\doteq J_1 n^{-1} + K_1 n^{-1}, & L_4 - L_4' &\doteq J_4 n^{-1} + K_4 n^{-2}, \\ L_2 - L_2' &\doteq J_2 n^{-1} + K_2 n^{-2}, & L_5 - L_5' &\doteq K_5 n^{-1}, \\ L_3 - L_3' &\doteq J_3 n^{-1} + K_3 n^{-1}, & L_6 - L_6' &\doteq K_6 n^{-2}. \end{aligned}\right\} \tag{2·23}$$

The J and K are the terms in the λ in (2·18). To n^{-2} (i.e. ignoring n^{-1}) the frequency generated from T of (2·19) is as follows:

$$\begin{aligned} f(t) = T &+ n^{-1}\left\{J_1 D + \frac{J_3}{6} D^3\right\} + n^{-1}\left\{\frac{D^2}{2}(J_2 + J_1^2) + \frac{D^4}{24}(J_4 + 4J_1 J_3) + \frac{D^6}{72} J_3^2\right\} \\ &+ n^{-1}\left\{K_1 D + \frac{D^3}{6}(K_3 + 3J_1 J_2 + J_1^3) + \frac{D^5}{120}(K_5 + 5J_1 J_4 + 10 J_2 J_3 + 10 J_1^2 J_3)\right. \\ &\left. + \frac{D^7}{144}(J_3 J_4 + 2J_1 J_3^2) + \frac{J_3^3}{1296} D^9\right\} + n^{-2}\left\{\frac{D^2}{2}(K_2 + 2J_1 K_1) + \frac{D^4}{24}(K_4 + 4J_1 K_3\right. \\ &+ 4J_3 K_1 + 3J_3^2 + 6J_1^2 J_2 + J_1^4) + \frac{D^6}{720}(K_6 + 6J_1 K_5 + 20 J_3 K_3 + 15 J_2 J_4 \\ &+ 60 J_1 J_2 J_3 + 15 J_1^2 J_4 + 20 J_1^3 J_3) + \frac{D^8}{11{,}520}(16 J_3 K_5 + 10 J_4^2 + 80 J_2 J_3^2 \\ &\left. + 80 J_1 J_3 J_4 + 80 J_1^2 J_3^2) + \frac{D^{10}}{5184}(3 J_3^2 J_4 + 4 J_1 J_3^3) + \frac{J_3^4}{31{,}104} D^{12}\right\}, \end{aligned} \tag{2·24}$$

with

$$D^h = \left(-\frac{d}{dt}\right)^h T.$$

To n^{-1}, (2·24) agrees with the formula given by M. S. Bartlett (1935), in which, however, there is a small and obvious slip in a sign. The law of formation of the numerical coefficients of (2·24) is evident; for instance, the numerical coefficient of $D^8 J_2 J_3^2$ is $1/144 = 1/2!\,3!^2 2!$. The integrals $\int_t^\infty$ and $\int_{-\infty}^{-t}$ $(t > 0)$ are found by reducing the exponent of D by unity, as follows:

$$\int_{-\infty}^{-t} D\,dt = -T, \quad \int_t^\infty D^{2m}\,dt = \int_{-\infty}^{-t} D^{2m}\,dt = D^{2m-1}, \quad \int_t^\infty D^{2m+1}\,dt = -\int_{-\infty}^{-t} D^{2m+1}\,dt = D^{2m}. \tag{2·25}$$

In normal theory the upper and lower $2\frac{1}{2}$ % points of t are $\pm 2{\cdot}262$ for $n = 10$. Table 2 shows the 'true' probabilities, i.e. the value of

$$\int_{-\infty}^{-2{\cdot}262} f(t)\,dt \tag{2·26}$$

for parent universes specified by λ_3, λ_4, using (2·24).

There are two observations to be made on the results presented in this table. The first is that, despite the considerable number of terms (shown at (2·24)) included in the probability expansion, the values found in the successive terms cannot be regarded as satisfactorily convergent for so small a sample as 10, and, of course, the convergence disimproves with increasing $\sqrt{\beta_1}$. Taken all together, however, they seem consistent and significant. The second observation is that attention was confined to the negative 'tail' of the distribution. It may be assumed that, in all cases, the distortion would be very considerably less marked if regard were had to the probability for $|t| > 2{\cdot}262$. Actually for universe 3 the probability

is 0·056, not significantly different from the normal theory probability of 0·05. In justification of the attitude adopted above, the point might be put as follows:

We decide to accept the hypothesis that the universal mean is zero provided that the value of t found from the particular sample satisfies $t_0 \leqslant t \leqslant t_1$, where

$$\text{Prob}\,(t < t_0) = \text{Prob}\,(t > t_1) = 0{\cdot}025.$$

The table is designed to show that if the parent universe is markedly asymmetrical the range (t_0, t_1) may differ appreciably from $-t_0 = t_1 = 2{\cdot}262$.

Table 2. *Probabilities of t less than* $-2{\cdot}262$ *for samples of* 10 *for seven universes*

Universe	$\lambda_3 = \sqrt{\beta_1}$	$\lambda_4 = \beta_2 - 3$	Probability
Normal	0	0	0·025
2	0	1	0·024
3	1/2	0	0·041
4	$1/\sqrt{2}$	1/2	0·047
5	1	0	0·072?
6	1	1	0·086?
7	1/2	1/2	0·043

As anticipated by earlier work (W. S. Gosset, 1908; R. C. Geary, 1936), the table shows that the distortion is slight for symmetrical universes; even when $\lambda_4 = 1$ (and $\lambda_3 = 0$) the probability (0·024) is practically identical with the normal value. There can be little doubt that the standard table probabilities can be seriously at variance with the true probabilities when the universes from which the samples are drawn are markedly asymmetrical.

(c) *Difference of means*

R. A. Fisher's (1925) test of significance

$$t = \frac{(k_1' - k_1'')\sqrt{(n' + n'' - 2)}}{\{(n'-1)\,k_2' + (n''-1)\,k_2''\}^{\frac{1}{2}}} \sqrt{\frac{n'n''}{(n'+n'')}}, \tag{2·27}$$

for the difference of averages k_1' and k_1'' in normal theory for random samples numbering n' and n'' is, of course, a particular case of the analysis of variance considered in § (*a*) above. The second cumulants are k_2' and k_2''. It is assumed that the unknown universal means and variances are equal. Suppose now that the random samples in reality have been derived from universes in which the means are equal but the other semi-invariants λ_i' and λ_i'' are not necessarily zero for $i \geqslant 2$, or even necessarily equal. Since the universal means are assumed equal, without loss of generality we may take $\lambda_1' = \lambda_1'' = 0$. This general mathematical model seems to be the correct one; we are not trying to determine the probability of the samples being derived from the *same universe* but rather if they could conceivably have been drawn from universes with the *same arithmetic mean*, however much they may differ otherwise. The correctness or otherwise of the concept may be considered in relation to, say, the problem of deciding from two random samples which of two types of fertilizer is to be preferred from yield observations on a given crop on a given kind of land. Undoubtedly the prime problem will be that of ascertaining which is probably the better yielding (i.e. whether the arithmetic means are significantly different). Of considerably less importance is the

question of which fertilizer is the more variable; of less importance still is the question of deciding, say, whether with approximately equal yields one universe is symmetrical and the other markedly asymmetrical. The point is that the question of the equality of universal means should be considered without assuming that the other semi-invariants in the universes from which the samples have been drawn are necessarily equal. This essentially is also the viewpoint in R. A. Fisher's randomization method.

Expanding the denominator of (2·27) in terms of $(k_2'-\lambda_2')$ and $(k_2''-\lambda_2'')$ and computing therefrom the first few terms of the first four moments of t, we find the following approximations to the first four semi-invariants:

$$\left.\begin{aligned}
AL_1 &\simeq -\frac{(\lambda_3'-\lambda_3'')}{2(n'\lambda_2'+n''\lambda_2'')},\\
A^2L_2 &\simeq \left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)\left(1+\frac{2\overline{n'\lambda_2'^2+n''\lambda_2''^2}}{n'\lambda_2'+n''\lambda_2''}\right)\\
&\qquad +\frac{(n'^2-n''^2)(\lambda_4'\lambda_2''-\lambda_4''\lambda_2')}{n'n''(n'\lambda_2'+n''\lambda_2'')^2}+\frac{7(\lambda_3'-\lambda_3'')^2}{4(n'\lambda_2'+n''\lambda_2'')^2},\\
A^3L_3 &\simeq \frac{\lambda_3'}{n'^2}-\frac{\lambda_3''}{n''^2}-\frac{3(\lambda_3'-\lambda_3'')}{(n'\lambda_2'+n''\lambda_2'')}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right),\\
A^4L_4 &\simeq \frac{6(n'\lambda_2'^2+n''\lambda_2''^2)}{(n'\lambda_2'+n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)^2-6\left(\frac{\lambda_3'}{n'^2}-\frac{\lambda_3''}{n''^2}\right)\frac{(\lambda_3'-\lambda_3'')}{(n'\lambda_2'+n''\lambda_2'')}\\
&\quad +\frac{18(\lambda_3'-\lambda_3'')^2}{(n'\lambda_2'+n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)+\frac{\lambda_4'}{n'^3}+\frac{\lambda_4''}{n''^3}\\
&\quad -3\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)\frac{\{\lambda_4'(n'n''\lambda_2'+2\overline{n''^2-n'^2}\lambda_2'')+\lambda_4''(n'n''\lambda_2''+2\overline{n'^2-n''^2}\lambda_2')\}}{n'n''(n'\lambda_2'+n''\lambda_2'')^2},
\end{aligned}\right\} \tag{2·28}$$

with
$$A=\left\{\left(\frac{n'+n''}{n'n''}\right)\frac{(\lambda_2'\overline{n'-1}+\lambda_2''\overline{n''-1})}{(n'+n''-2)}\right\}^{\frac{1}{2}}.$$

Using formula (2·24) to the term in n^{-1} with the Gosset-Fisher function again as generating function, Table 3 shows rough approximations, for four examples, to the 'true' probability of values of $t\leqslant\tau$, where τ is the (negative) value for probability 0·025 from the normal table, and $\lambda_2'=\lambda_2''=1$. When the two samples are drawn from different universes the distortion can accordingly be considerable. The third example suggests that if the universes are the same the distortion is small, a result to be anticipated from the fact (apparent from (2·28)) that, to the approximation used, the first two semi-invariants are equal to their normal theory values; this theory confirms the experimental results of E. S. Pearson & N. K. Adyanthaya (1929).

Table 3

Example	n'	n''	λ_3'	λ_3''	λ_4'	λ_4''	Probability
1	12	4	1	−1	1	−1	0·045
2	18	6	1	−1	1	−1	0·041
3	7	4	$1/\sqrt{2}$	$1/\sqrt{2}$	1/2	1/2	0·027
4	10	6	1	0	1	0	0·036

It should be remarked that the probabilities in Table 3 (as well as in Table 2) are merely rough approximations—the samples used are far too small for the results to have any pretension to accuracy. The object has been merely to show that the actual probability *could* be considerably at variance with that shown in the standard table, for small samples.

3. Sufficient conditions for approach to normality of $a(c)$ with increasing n

The remainder of the paper deals with the field of symmetrical tests of normality, homogeneous of degree zero, represented by (3·1). It is essential to establish the conditions of approach to normality of the frequency distribution of $a(c)$ as the sample number increases.

Let
$$a(c) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^c \Big/ \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}c}, \tag{3·1}$$

where $\bar{x} = \Sigma x_i/n$ and c is non-negative. It will be shown in succession that, subject to stated conditions, with increasing n,

(i) the frequency distribution of
$$a_1(c) = \frac{1}{n}\Sigma |x_i|^c \Big/ \left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c} \tag{3·2}$$
tends towards normality, and

(ii) the frequency distribution of $a_1(c)$ tends towards that of $a(c)$ and hence towards normality.

It is assumed, without loss of generality, that the universal mean of the universe from which the sample of n is drawn is zero. Denote the kth absolute moment from zero by $\mu_{|k|}$, k not being necessarily an integer. Given a positive quantity ϵ arbitrarily small, $\omega(\epsilon)$ can be found so that
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|x_i|^c - \mu_{|c|})\right| < \omega\sqrt{\frac{(\mu_{|2c|}-\mu_{|c|}^2)}{n}}\right\} > 1-\epsilon, \tag{3·3}$$

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(x_i^2 - \mu_2)\right| < \omega\sqrt{\frac{(\mu_4-\mu_2^2)}{n}}\right\} > 1-\epsilon, \tag{3·4}$$

provided, of course, that $\mu_{|2c|}$ and μ_4 exist. As n increases ω may be envisaged as approaching the normal probability point appropriate to the probability ϵ, since, in the conditions stated, $\Sigma|x_i|^c/n$ and $\Sigma x_i^2/n$ are normally distributed in the limit. For samples which satisfy the inequality in the brackets { } at (3·4) and if n is so large that
$$\omega\sqrt{\left(\frac{\mu_4}{n}\right)} < \mu_2,$$
the denominator of (3·2) can be expanded to three terms (including the remainder) by Taylor's theorem, so that $a_1(c)$ may be written

$$a_1(c) = \mu_{|c|}\mu_2^{-\frac{1}{2}c}\left\{1 + \frac{1}{n}\Sigma\left(y_i - \frac{c}{2}z_i\right) - \frac{c}{2n^2}\Sigma y_i\Sigma z_i + \frac{c(c+2)}{8n^2}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X\right\}, \tag{3·5}$$

with
$$y_i = (|x_i|^c - \mu_{|c|})/\mu_{|c|},$$
$$z_i = (x_i^2 - \mu_2)/\mu_2,$$
$$X = \mu_2^{\frac{1}{2}c+2}\left\{\mu_2 + \frac{\theta}{n}\Sigma(x_i^2-\mu_2)\right\}^{-\frac{1}{2}(c+4)} \quad (0<\theta<1).$$

With probability exceeding $(1-\epsilon)$ it is evident, from (3·4), that X is maximized by

$$\left(1-\frac{\omega}{\mu_2\sqrt{\frac{\mu_4}{n}}}\right)^{-\frac{1}{2}(c+4)}$$

It will suffice, for the present purpose, to infer that

$$|X|<\kappa,$$

where κ is a constant independent of n. We have now

$$E\frac{1}{n}\Sigma\left(y_i-\frac{c}{2}z_i\right)=0.$$

Set
$$\sigma^2=E\frac{1}{n^2}\left\{\Sigma\left(y_i-\frac{c}{2}z_i\right)\right\}^2$$

$$=\frac{1}{n}\left\{\frac{\mu_{|2c|}}{\mu_{|c|}^2}-\frac{c\mu_{|c+2|}}{\mu_{|c|}\mu_2}+\frac{c^2}{4}\frac{\mu_4}{\mu_2^2}-\left(\frac{c}{2}-1\right)^2\right\}, \tag{3·6}$$

and
$$\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right)-\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right)=u, \tag{3·7}$$

with
$$u=-\frac{c}{2n^2\sigma}\Sigma y_i\,\Sigma z_i+\frac{c(c+2)}{8n^2\sigma}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X. \tag{3·8}$$

For samples which satisfy the inequalities in { } at (3·3) and (3·4) and hence with a probability exceeding $(1-2\epsilon)$, we have

$$|u|<\frac{c\omega^2}{2\sigma}\frac{\sqrt{(\mu_{|2c|}\mu_4)}}{n\mu_{|c|}\mu_2}+\frac{c(c+2)}{8\sigma}\frac{\kappa\omega^2\mu_4}{n\mu_2^2}\left(1+\frac{\omega}{\mu_{|c|}}\sqrt{\frac{\mu_{|2c|}}{n}}\right)<\frac{\xi}{\sqrt{n}}, \tag{3·9}$$

where ξ is independent of n. Or, briefly,

$$\text{Prob}\left\{|u|<\frac{\xi}{\sqrt{n}}\right\}>1-2\epsilon, \tag{3·10}$$

so that u tends in probability towards zero with $1/n$. Now (3·7) may be written in the form $u=Y'-Y$, where Y' and Y are the respective terms on the left side. If A be any number and F the total probability function, a well-known lemma (Fréchet, 1937, p. 164) shows that

$$|F_{Y'}(A)-F_Y(A)|\leqslant\left\{F_Y\left(A+\frac{\xi}{\sqrt{n}}\right)-F_Y\left(A-\frac{\xi}{\sqrt{n}}\right)\right\}+2\epsilon, \tag{3·11}$$

using (3·10). Hence the frequency distribution of

$$Y'=\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right) \tag{3·12}$$

tends towards that of
$$Y=\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right) \tag{3·13}$$

at every continuity point of the latter frequency, as n tends towards infinity. But Y, from (3·13), is the simple average of n random measures, and its frequency must tend towards normality provided that its standard deviation exists; from (3·6) it is evident that σ is finite provided that $\mu_{|k|}$, where k is the greater of $2c$ and 4, is finite. Here and in the remainder of this section it will be useful to remember that if $\mu_{|k|}$ exists so does $\mu_{|k'|}$ for $0\leqslant k'\leqslant k$.

To prove that the frequency distribution of $a(c)$ tends towards that of $a_1(c)$ and hence towards normality with increasing n it will be shown that $\Sigma|x_i-\bar{x}|^c/n$ tends in probability towards $\Sigma|x_i|^c/n$. Two cases will be considered separately: (1) $c \geqslant 1$, (2) $1 > c \geqslant 0$.

Case (1). $c \geqslant 1$

For values of x_i for which $|x_i| \geqslant |\bar{x}|$,

$$|x_i-\bar{x}|^c-|x_i|^c = \pm c\bar{x}\,|x_i-\theta\bar{x}|^{c-1} \quad (0<\theta<1)$$

and when $|x_i| < |\bar{x}|$, $\quad ||x_i-\bar{x}|^c-|x_i|^c| \leqq (2^c+1)\,|\bar{x}|^c$.

Hence
$$\frac{1}{n}\left|\sum_{i=1}^{n}(|x_i-\bar{x}|)^c-(|x_i|^c)\right| < |\bar{x}|\left(\frac{B}{n}\sum_{i=1}^{n}|x_i|^{c-1}+C\,|\bar{x}|^{c-1}\right), \tag{3·14}$$

B and C being independent of the x_i and n but depending on c. With ϵ arbitrarily small ω can be found so that

$$\left.\begin{aligned}&\text{Prob}\left\{|\bar{x}| < \omega\sqrt{\frac{\mu_2}{n}}\right\} > 1-\epsilon,\\ &\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|x_i|^{c-1}-\mu_{|c-1|})\right| < \omega\sqrt{\frac{\mu_{|2c-2|}-\mu^2_{|c-1|}}{n}}\right\} > 1-\epsilon.\end{aligned}\right\} \tag{3·15}$$

Hence, from (3·14) and (3·15), if μ_2 and $\mu_{|2c-2|}$ exist,

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma|x_i-\bar{x}|^c-\frac{1}{n}\Sigma|x_i|^c\right| < B'\frac{\omega\mu_{|c-1|}\mu_2^{\frac{1}{2}}}{\sqrt{n}}\right\} > 1-2\epsilon$$

for n sufficiently large the constant B' depending on c but not on n. Hence for $c \geqq 1$, $\Sigma|x_i-\bar{x}|^c/n$ tends in probability towards $\Sigma|x_i|^c/n$. Incidentally, this proves that $\{\Sigma(x_i-\bar{x})^2/n\}^{\frac{1}{2}c}$ tends in probability towards $\{\Sigma x_i^2/n\}^{\frac{1}{2}c}$, the latter two expressions representing respectively the denominators of $a(c)$ and $a_1(c)$.

Case (2). $1 > c \geqslant 0$

Let $\bar{x}$ satisfy a probabilistic inequality identical in form with the first equation of (3·15) and let γ be any positive quantity, fixed once for all. Let n (presently to be defined further) be so large that

$$\gamma > \omega\sqrt{\frac{\mu_2}{n}}.$$

Then
$$\frac{1}{n}\sum_{i=1}^{n}(|x_i-\bar{x}|^c-|x_i|^c) = \frac{1}{n}\left(\sum_{|x_i|\geqslant\gamma}' + \sum_{|x_i|<\gamma}''\right)(|x_i-\bar{x}|^c-|x_i|^c). \tag{3·16}$$

When $|x_i| \geqslant \gamma$ (i.e. in Σ'),

$$|x_i-\bar{x}|^c-|x_i|^c = \pm c\bar{x}\,|x_i-\theta\bar{x}|^{c-1} \quad (0<\theta<1),$$

so that
$$\text{Prob}\left\{||x_i-\bar{x}|^c-|x_i|^c| < c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}\right\} > 1-\epsilon. \tag{3·17}$$

When $|x_i| < \gamma$ (i.e. in Σ''), given η arbitrarily small and positive, n can be found so that

$$||x_i-\bar{x}|^c-|x_i|^c| < \eta, \tag{3·18}$$

when
$$|\bar{x}| < \omega\sqrt{\frac{\mu_2}{n}},$$

since $|x|^c$ $(c>0)$ is uniformly continuous in Σ''. We then have

$$\text{Prob}\{||x_i-\bar{x}|^c-|x_i|^c| < \eta\} > 1-\epsilon. \tag{3·19}$$

Combining (3·17) and (3·19), it may be inferred that

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma|x_i-\bar{x}|^c-\frac{1}{n}\Sigma|x_i|^c\right|<c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}+\eta\right\}>1-2\varepsilon, \qquad (3\cdot20)$$

the first term of the upper limit in { } tending to zero as n tends towards infinity, and ϵ and η being arbitrarily small

We have accordingly shown that the numerator and denominator of $a(c)$ tends in probability towards those of $a_1(c)$. Hence $a(c)$ tends in probability towards $a_1(c)$. Hence, using the lemma cited at (3·11), the total frequency of $a(c)$ tends towards that of $a_1(c)$ which tends towards normality as n tends towards infinity. Finally:

If $c \geqslant 0$ the frequency distribution of $a(c)$, given by (3·1), tends towards normality as n tends towards infinity provided that $\mu_{|k|}$, where k is the greater of $2c$ and 4, is finite.

It seems likely that an analogous theorem can be proved for $0 > c > -\frac{1}{2}$; we shall not, however, be concerned in this communication with negative values of c.

4. Moments of $a(c)$ for normal samples

While it will be shown in later sections that, with indefinitely large samples, $\sqrt{b_1}$ and b_2 are the most efficient tests of asymmetry and kurtosis, respectively, it by no means follows that other tests are inefficient or that they may not be useful supplements in cases in which the prime tests are indecisive as to the probable non-normality of a given sample. It is accordingly proposed to give here close approximations to the first four moments (from the origin) of $a(c)$ (given by (3·1)) for normal random samples of n.

For normal samples (R. A. Fisher, 1929; R. C. Geary, 1933)

$$M'_k\{a(c)\} = E\{a(c)\}^k = E\left\{\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|^c\right\}^k \Big/ E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}ck} \qquad (4\cdot1)$$

The exact value of the denominator is, of course, known, for

$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}k'} = \left(\frac{n-1}{n}\right)^{\frac{1}{2}k'} Es^{k'} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-3}{2}\right)!\Big/\left(\frac{n-3}{2}\right)!, \qquad (4\cdot2)$$

since, as usual, $(n-1)s^2 = \Sigma(x_i-\bar{x})^2$. It will be useful to expand $\log_e Es^{k'}$ with $k' = ck$ using Stirling's formula in (4·2):

$$\begin{aligned}\log_e Es^{k'} &= \frac{k'}{2}\log\frac{2}{n-1}+\log\left(\frac{n+k'-3}{2}\right)!-\log\left(\frac{n-3}{2}\right)! \\ &= \frac{(k'^2-2k')}{4(n-1)}-\frac{k'(k'-1)(k'-2)}{12(n-1)^2}+\frac{k'^2(k'-2)^2}{24(n-1)^3}-\frac{k'(k'-1)(k'-2)(3k'^2-6k'-4)}{120(n-1)^4} \\ &\quad+\frac{k'^2(k'-2)^2(k'^2-2k'-2)}{60(n-1)^5}-\frac{k'(k'-1)(k'-2)(3k'^4-12k'^3+24k'+16)}{252(n-1)^6} \\ &\quad+\frac{k'^2(k'-2)^2(3k'^4-12k'^3-4k'^2+32k'+32)}{336(n-1)^7}, \qquad (4\cdot3)\end{aligned}$$

which checks for $k' = 1$ to $(n-1)^{-7}$ with Geary (1935, p. 354). Take

$$v(c) = \frac{1}{n}\sum_{i=1}^{n}|z_i|^c, \qquad (4\cdot4)$$

with

$$z_i = x_i - \bar{x}.$$

The moments of $v(c)$ will be found exactly as in the case of $c = 1$ (Geary, 1936) from the single or joint normal frequency distributions of $(z_1, z_2, \ldots)$. We find

$$M'_1\{v(c)\} = \frac{1}{\sqrt{\pi}}\left(\frac{2\overline{n-1}}{n}\right)^{\frac{1}{2}c}\left(\frac{c-1}{2}\right)!,$$

$$M'_2\{v(c)\} = \frac{1}{\sqrt{\pi}}\frac{(2\overline{n-1})^c}{n^{c+1}}\left(\frac{2c-1}{2}\right)! + \frac{2^c}{\pi}n^{-\frac{1}{2}}(n-1)^{-c}(n-2)^{\frac{1}{2}(2c+1)}\left[\left(\frac{c-1}{2}\right)!\right]^2 \tag{4·5}$$

$$\times\left\{1+\frac{1}{2!}\left(\frac{c+1}{2}\right)^2\left(\frac{2}{n-1}\right)^2+\frac{1}{4!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{2}{n-1}\right)^4+\frac{1}{6!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{c+5}{2}\right)^2\left(\frac{2}{n-1}\right)^6+\ldots\right\}. \tag{4·6}$$

For the third moment we write

$$M'_3\{v(c)\} = E\{v(c)\}^3 = \frac{n}{n^3}\Sigma|z_1|^{3c} + \frac{3n(n-1)}{n^3}E|z_1|^{2c}|z_2|^c + \frac{n(n-1)(n-2)}{n^3}E|z_1|^c|z_2|^c|z_3|^c$$

$$= A_1 + A_2 + A_3, \tag{4·7}$$

denoting the three terms on the right by A_1, A_2, A_3 respectively. Then

$$A_1 = \frac{1}{\sqrt{\pi}}\left(\frac{3c-1}{2}\right)!(2\overline{n-1})^{\frac{1}{2}c}n^{-\frac{1}{2}(4+3c)},$$

$$A_2 = \frac{3.2^{\frac{1}{2}c}}{\pi}\left(\frac{2c-1}{2}\right)!\left(\frac{c-1}{2}\right)!(n-2)^{\frac{1}{2}(3c+1)}(n-1)^{-\frac{1}{2}c}n^{-\frac{1}{2}}$$

$$\times\left\{1+\frac{(2c+1)(c+1)}{2!(n-1)^2}+\frac{(2c+3)(2c+1)(c+3)(c+1)}{4!(n-1)^4}+\ldots\right\}.$$

$$A_3 = \left(\frac{2^c}{\pi}\right)^{\frac{1}{2}}(n-3)^{\frac{1}{2}(3c+2)}(n-2)^{-\frac{1}{2}(3c+1)}(n-1)n^{-\frac{1}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^3\left\{1+\frac{3}{2}\frac{(c+1)^2}{(n-2)^2}\right.$$

$$-\frac{(c+1)^3}{(n-2)^3}+\frac{(c+1)^2(c+3)(7c+9)}{8(n-2)^4}-\frac{(c+3)^2(c+1)^3}{2(n-2)^5}$$

$$\left.+\frac{(c+3)^2(c+1)^2(61c^2+310c+265)}{240(n-2)^6}+\ldots\right\}.$$

Similarly, for the fourth moment,

$$M'_4\{v(c)\} = E\{v(c)\}^4 = \frac{n}{n^4}E|z_1|^{4c} + \frac{4n(n-1)}{n^4}E|z_1|^{3c}|z_2|^c$$

$$+\frac{3n(n-1)}{n^4}E|z_1|^{2c}|z_2|^{2c} + \frac{6n(n-1)(n-2)}{n^4}E|z_1|^{2c}|z_2|^c|z_3|^c$$

$$+\frac{n(n-1)(n-2)(n-3)}{n^4}E|z_1|^c|z_2|^c|z_3|^c|z_4|^c$$

$$= C_1 + C_2 + C_3 + C_4 + C_5 \tag{4·8}$$

with

$$C_1 = \frac{2^{2c}}{\sqrt{\pi}}(n-1)^{2c}n^{-2c-\frac{1}{2}}\left(\frac{4c-1}{2}\right)!,$$

$$C_2 = \frac{2^{2(c+1)}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{1}{2}}\left(\frac{3c-1}{2}\right)!\left(\frac{c-1}{2}\right)!$$

$$\times\left\{1+\frac{(3c+1)(c+1)}{2!(n-1)^2}+\frac{(3c+3)(3c+1)(c+3)(c+1)}{4!(n-1)^4}+\ldots\right\},$$

$$C_3 = \frac{3.2^{2c}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{1}{2}}\left[\left\{\left(\frac{2c-1}{2}\right)!\right\}^2\left\{1+\frac{(2c+1)^2}{2!(n-1)^2}+\frac{(2c+3)^2(2c+1)^2}{4!(n-1)^4}+\ldots\right\}\right.,$$

$$C_4 = \frac{3.2^{2c+1}}{\pi^{\frac{1}{2}}}(n-3)^{2c+1}(n-2)^{-\frac{1}{2}(4c+1)}(n-1)n^{-\frac{1}{2}}\left(\frac{2c-1}{2}\right)!\left[\left\{\left(\frac{c-1}{2}\right)\right\}!\right]^2$$

$$\times\left\{1+\frac{(c+1)(5c+3)}{2(n-2)^2}-\frac{(c+1)^2(2c+1)}{(n-2)^3}+\frac{(c+1)(57c^3+227c^2+255c+81)}{24(n-2)^4}\right.$$

$$\left.-\frac{(2c+1)(c+1)^2(c+3)(5c+9)}{6(n-2)^5}+\ldots\right\},$$

$$C_5 = \frac{2^{2c}}{\pi^2}(n-4)^{\frac{1}{2}(4c+5)}(n-3)^{-2c-1}(n-2)(n-1)n^{-\frac{1}{2}}\left[\left\{\left(\frac{c-1}{2}\right)!\right\}^4\right.$$

$$+\left\{1+\frac{3(c+1)^2}{(n-3)^2}-\frac{4(c+1)^3}{(n-3)^3}+\frac{(c+1)^2(7c^2+21c+15)}{(n-3)^4}-\frac{4(c+3)(c+1)^3(2c+3)}{(n-3)^5}\right.$$

$$\left.+\frac{(c+3)(c+1)^2(122c^3+671c^2+1070c+525)}{15(n-3)^6}-\ldots\right\}.$$

Formulae (4·5), (4·6), (4·7) and (4·8) were checked from the corresponding formulae for $c = 1$ given in the author's 1936 paper.

From the following section it will be apparent that for indefinitely large samples the most sensitive test of kurtosis of the field $a(c)$ is found for $c = 4$. At the same time it is shown that there is really not much difference in efficiency for values of c in the range $5 \geqslant c > 2$; moreover, the results in § 6 (in which the efficiency of the tests for $c = 4$ and $c = 1$ are compared from the power function viewpoint) suggest that, for samples of moderate size, the superiority, if any at all, of a test using $a(4) = b_2$ over other tests in the series may be even less marked. The disadvantage of $a(4)$ is that its frequency is not known for samples of all sizes; and if we could estimate, with any degree of confidence, the probability points of $a(c)$ for any value or values of $c > 2$ for medium-size samples we might, for practical purposes, dispense with $a(4)$ altogether, since, while we now know one way of solving the problem of determining the exact, or almost exact, frequency distribution of $a(4)$, it must be admitted that the method is extremely tedious. (From the theoretical point of view, however, the $a(4)$ problem must be solved since it remains a challenge to the mathematical skill of statisticians!) It will accordingly be of interest to study the order of magnitude of the semi-invariants of $a(c)$ for c near 2.

Consider the case, for example, of $c = 2{\cdot}4$, not by any means, it is important to observe, the lowest value which would be used for tabulating. In Table 4 the first three moments are given for $n = 25$. The L's represent, of course, the semi-invariants. The values of the functions for $a_1(c)$ (given by (3·2)) for $n = 24$ (i.e. the appropriate number of degrees of freedom for comparison with $a(c)$) are also given. These show that the moments of $a_1(c)$ are very close to those of $a(c)$, which suggests that, when n is not less than, say, 20, the values of B_1, B_2 and corresponding functions of higher orders, if required, for $a_1(c)$ could be used for the determination of the probability points of $a(c)$. This is important from the computational point of view because the algebraic expressions for the normal moments of $a_1(c)$ are exceedingly simple whereas it must be conceded that (4·8) offers a grim prospect for the computer; furthermore, the principal term C_5 is rather slowly convergent unless $n > 50$ or so,

whereas *exact* values for all values of n can readily be found for the moments of $a_1(c)$ for normal samples.

Table 4. *Normal moments, etc., of* $a(c)$ *and* $a_1(c)$ *for* $c = 2{\cdot}4$

	$a(2{\cdot}4)$	$a_1(2{\cdot}4)$
n	25	24
$M_1' = L_1$	1·166252	1·1662524891
M_2'	1·362004	1·362091186
M_3'	1·592841	1·593151615
$M_2 = L_2$	0·001860	0·001946318
$M_3 = L_3$	0·000063	0·000069583
$\sqrt{B_1} = L_3/L_2^{\frac{3}{2}}$	0·80	0·8104

As with (4·1) for $a(c)$, the moments (from the origin) of any order of $a_1(c)$ is the quotient of the moments of the same order for numerator and denominator, assuming that the universal mean is zero and the variance unity. Since the different members x_i of the sample are independent—the difficulty with $a(c)$ is that the $(x_i - \bar{x})$ are *not* independent—for the moments of the numerator of (3·2) we require only

$$E\,|x|^{k'} = \frac{2}{\sqrt{(2\pi)}}\int_0^\infty dx\, x^{k'} e^{-\frac{1}{2}x^2} = \left(\frac{k'-1}{2}\right)!\,\frac{2^{\frac{1}{2}k'}}{\sqrt{\pi}}, \tag{4·9}$$

and for the denominator

$$Es^{k'} = E\left(\frac{1}{n}\sum_i x_i^2\right)^{\frac{1}{2}k'} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-2}{2}\right)!\Big/\left(\frac{n-2}{2}\right)!. \tag{4·10}$$

The case of $c = 4$ is particularly simple. The first four semi-invariants are as follows:

$$\left.\begin{aligned} L_1 &= M_1' = \frac{3n}{(n+2)},\\ L_2 &= M_2 = \frac{24n^2(n-1)}{(n+2)^2(n+4)(n+6)},\\ L_3 &= M_3 = \frac{1728(n-1)(n-2)n^3}{(n+2)^3(n+4)(n+6)(n+8)(n+10)},\\ L_4 &= \frac{10{,}368n^4(n-1)(30n^4+168n^3-608n^2-2672n+3712)}{(n+2)^4(n+4)^2(n+6)^2(n+8)(n+10)(n+12)(n+14)}. \end{aligned}\right\} \tag{4·11}$$

Moments, etc., for $a_1(c)$ for normal samples of 24 and 50 are contrasted for $c = 2{\cdot}4$ and $c = 4$ in Table 5. The contrast between the values of $\sqrt{B_1}$ and $(B_2 - 3)$ respectively for $a_1(2{\cdot}4)$ and $a_1(4)$ is striking in the extreme. Even for $n = 24$ $\sqrt{B_1}[a_1(2{\cdot}4)]$ and $B_2[a_1(2{\cdot}4)]$ are approaching the values at which a Gram-Charlier approximation to the frequency distribution may be reasonably convergent. Furthermore, the decline in the values of the B's from $n = 24$ to $n = 50$ is marked for $a_1(2{\cdot}4)$, while the decline in the $B[a_1(4)]$ is very slow.

It is accordingly suggested that a table of probability points (perhaps 0·001, 0·01, 0·025, 0·05 and 0·10) of $a(c)$, for c equal to, say, 2·2, be prepared for $n \geqslant 25$ on the assumption that Gram-Charlier applies throughout. For this purpose the values of the mean and variance for n at intervals of, say, 10 should be computed from formulae (4·5) and (4·6); the B_1 and $(B_2 - 3)$ should, however, be computed as for $a_1(c)$. For lower sample sizes it might be well

to use terms to order n^{-2} which would render necessary the use of the fifth and sixth semi-invariants of $a_1(c)$. The formulae given by E. A. Cornish & R. A. Fisher (1937) (assuming Gram-Charlier) could be used to find the probability points. On account of the minuteness of the variance L_2 for c near 2 it will be necessary to work to many places of decimals—at least 10. As stated at the outset, the test of kurtosis $a(2{\cdot}2)$ will be only slightly less efficient than $a(4)$ and it may be slightly more efficient than $a(1)$, the probability points of which are known approximately for samples of all sizes. In any case the $a(2{\cdot}2)$ table would be a useful adjunct to that of $a(1)$.

Table 5. *Normal moments, etc., of* $a_1(c)$ *for* $c = 2{\cdot}4$ *and* $c = 4$

	$n=24$		$n=50$	
	$c=2{\cdot}4$	$c=4$	$c=2{\cdot}4$	$c=4$
$M_1' = L_1$	1·1662524891	2·769231	1·1721603127	2·884615
$M_2 = L_2$	0·001946318	0·559932	0·001058462	0·359550
$M_3 = L_3$	0·000069583	0·752488	0·000022251	0·343337
$L_4 = M_4 - 3L_2^2$	0·000004921	1·955999	0·000000919	0·711375
$\sqrt{B_1} = L_3/L_2^{\frac{3}{2}}$	0·8104	1·7960	0·6462	1·5925
$B_2 - 3 = L_4/L_2^2$	1·30	6·24	0·82	5·50

In an earlier paper (1935) the writer suggested that the correlation between b_2 and $a(1)$ for normal samples gave some indication of the relative efficiency of these two tests of normality. In this order of ideas it seems desirable to compute the approximate value of the correlation coefficient between $a(c)$ and $a(c')$, where c and c' are any two positive constants. In the first instance the universe from which the sample of n was drawn was not necessarily normal. Since in the present application we will be concerned only with large samples we assume the universal mean known (and accordingly it may be taken as zero, i.e. $\lambda_1 = 0$), so that, instead of $a(c)$ we use, in reality, $a_1(c)$ given by (3·2). In the remainder of this section we write a for $a_1(c)$ and a' for $a_1(c')$:

$$a = \left(\frac{1}{n}\Sigma\,|x_i|^c\right) \Big/ \left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c}, \tag{4·12}$$

$$a' = \left(\frac{1}{n}\Sigma\,|x_i|^{c'}\right) \Big/ \left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c'}. \tag{4·13}$$

Set

$$\left.\begin{aligned} y_i &= (|x_i|^c - \mu_{|c|})/\mu_{|c|}, \\ y_i' &= (|x_i|^{c'} - \mu_{|c'|})/\mu_{|c'|}, \\ z_i &= (x_i^2 - \mu_2)/\mu_2, \\ \alpha &= \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad \alpha' = \mu_{|c'|}/\mu_2^{\frac{1}{2}c'}, \\ C &= \frac{c+c'}{2}, \quad C_k = \frac{C(C+1)(C+2)\ldots(C+k-1)}{k!}. \end{aligned}\right\} \tag{4·14}$$

Then

$$\frac{aa'}{\alpha\alpha'} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma y_i'\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}(c+c')} \tag{4·15}$$

The mean value of $aa'/\alpha\alpha'$ was found approximately (i.e. to terms in n^{-3}) by formally expanding the last factor in (4·15), multiplying by the first two factors, and setting down the mean value term by term, so that

$$\begin{aligned}
M'_{cc'}/\alpha\alpha' = Eaa'/\alpha\alpha' \simeq \Big\{1 &+ \frac{1}{n^2}C_2 nEz^2 - \frac{1}{n^3}C_3 nEz^3 \\
&+ \frac{1}{n^4}C_4\left(nEz^4 + \frac{6n\overline{n-1}}{2}E^2z^2\right) - \frac{1}{n^5}C_5(10n\overline{n-1}\,Ez^3Ez^2) \\
&+ \frac{1}{n^6}C_6 90\frac{n\overline{n-1}\,\overline{n-2}}{6}E^3z^2\Big\} + \Big\{-\frac{C_1}{n^2}(nEyz + nEy'z) + \frac{C_2}{n^3}(nEyz^2 + nEy'z^2) \\
&- \frac{C_3}{n^4}[n(Eyz^3 + Ey'z^3) + 3n\overline{n-1}\,Ez^2(Eyz + Ey'z)] \\
&+ \frac{C_4}{n^5}[4n\overline{n-1}\,Ez^3(Eyz + Ey'z) + 6n\overline{n-1}\,Ez^2(Eyz^2 + Ey'z^2)] \\
&- \frac{30C_5}{n^6}\frac{n\overline{n-1}\,\overline{n-2}}{2}E^2z^2(Eyz + Ey'z)\Big\} + \Big\{\frac{1}{n^2}nEyy' - \frac{C_1}{n^3}nEyy'z \\
&+ \frac{C_2}{n^4}[nEyy'z^2 + 2n\overline{n-1}\,Eyz\,Ey'z + n\overline{n-1}\,Eyy'\,Ez^2] \\
&- \frac{C_3}{n^5}[n\overline{n-1}\,Eyy'\,Ez^3 + 3n\overline{n-1}\,(Eyz^2Ey'z + Ey'z^2Eyz + Eyy'zEz^2)] \\
&+ \frac{C_4}{n^6}\left[\frac{6n\overline{n-1}\,\overline{n-2}}{2}Eyy'E^2z^2 + 12n\overline{n-1}\,\overline{n-2}\,EyzEy'zEz^2\right]\Big\}. \qquad (4\cdot16)
\end{aligned}$$

The E's in (4·16) are readily calculable from (4·14), e.g.

$$Eyy' = Ey_iy'_i = E(|x_i|^c - \mu_{|c|})(|x_i|^{c'} - \mu_{|c'|})/\mu_{|c|}\mu_{|c'|} = (\mu_{|c+c'|}/\mu_{|c|}\mu_{|c'|}) - 1.$$

It has been verified that when c is substituted for c' in (4·13) the formula agrees with that for the second moment of $a_1(c)$ given in § 6.

The coefficient of correlation is, of course,

$$R_{cc'} = M_{cc'}/\sqrt{(M_{cc}M_{c'c'})}, \qquad (4\cdot17)$$

with $$M_{cc'} = M'_{cc'} - M'_cM'_{c'}.$$

Formulae for the first and second moments, to the approximation required, for the computation of (4·17) are given in § 6.

As an application, the following are the values of the variances and the covariance for the test of normality $a(1)$ and (b_2), i.e. in which c and c' have respectively the values 1 and 4, and where the universe belongs to the Pearson system with $\lambda_2 = 1$, $\lambda_3 = 0$ and $\lambda_4 = \frac{1}{2}$:

$$\left.\begin{aligned}
\frac{M_{cc}}{\mu^2_{|c|}} &\simeq \frac{0\cdot09313705}{n} - \frac{0\cdot262961}{n^2} - \frac{0\cdot196477}{n^3}, \\
\frac{M_{c'c'}}{\mu^2_{|c'|}} &\simeq \frac{4\cdot4286}{n} - \frac{92\cdot25}{n^2} + \frac{831\cdot2}{n^3}, \\
\frac{M_{cc'}}{\mu_{|c|}\mu_{|c'|}} &\simeq -\frac{0\cdot491}{n} + \frac{4\cdot87}{n^2} - \frac{281\cdot5}{n^3}.
\end{aligned}\right\} \qquad (4\cdot18)$$

From (4·17) and (4·18), $R_{cc'}(n=100) \doteq -0{\cdot}826$ and $R_{cc'}(n=\infty) = -0{\cdot}764$. It is of great interest to find that, though the universe is markedly non-normal the correlation for indefinitely large samples is practically identical with the normal theory value of $-0{\cdot}767$ (Geary, 1935), another indication, no doubt, that normal theory inferences can usually be applied with confidence when the parent universe is not markedly unsymmetrical.

When samples are indefinitely large we find, from (4·16) and (4·17),

$$R_{cc'} = \frac{4\mu_{|c+c'|} - 2(c\mu_{|c|}\mu_{|c'+2|} + c'\mu_{|c'|}\mu_{|c+2|}) + (cc'\mu_4 - \overline{c-2}.\overline{c'-2})\mu_{|c|}\mu_{|c'|}}{\sqrt{(M_{cc}M_{c'c'})}}, \qquad (4{\cdot}19)$$

where, of course, the values to be taken here for M_{cc} and $M_{c'c'}$ are found by substituting respectively c' for c and c for c' in the numerator. When, in addition, the parent universe is normal, we find

$$R^0_{cc'} = \frac{\left|\left(\frac{c+c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}\right)!\left(\frac{c'-1}{2}\right)!\left(\frac{cc'+2}{2}\right)\right|}{\sqrt{\left[\left\{\left(\frac{2c-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}!\right)^2\left(\frac{c^2+2}{2}\right)\right\}\left\{\left(\frac{2c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c'-1}{2}!\right)^2\left(\frac{c'^2+2}{2}\right)\right\}\right]}} \qquad (4{\cdot}20)$$

which reduces to $-1/\sqrt{\{12(\pi-3)\}}$ for $c=1$, $c'=4$, as it should (Geary, 1935). The following section will accord b_2 (i.e. $a(4)$) a decided primacy amongst tests of normality when the samples are indefinitely large. It may, therefore, be of interest to give the values of the correlation coefficients (for indefinitely large normal samples) between b_2 and $a(c)$ for selected values of c (Table 6). The table suggests, in the high coefficients of correlation, except for c very near 0 or 2, that all the $a(c)$ should be reliable tests of kurtosis, with no great difference between their efficiencies. The efficiency of any two tests would be identical, in the conditions stated, if the coefficient of correlation between them was ± 1 because then, of course, they would be functionally, and not stochastically, related.

Table 6. *Correlation between b_2 and $a(c)$ for indefinitely large normal samples*

Value of c	Value of R^0_{c4}	Value of c	Value of R^0_{c4}
0	0	3	0·980
1	−0·769	4	1
2	0	5	0·983
2·2	0·887	6	0·939
2·5	0·952	∞	0

5. The most efficient tests for indefinitely large samples

In this section we consider the efficiency of tests of kurtosis and asymmetry from the viewpoint of indefinitely large samples.

By definition a test will be regarded as *valid*, in relation to a field of continuous alternative universes including the normal, if its value for infinite samples drawn at random from the normal universe is different from its value for infinite samples from other universes of the field. As the sample number increases the test will become increasingly discriminatory of the normal as distinct from other universes of the field. This increased sensitivity might be given mathematical expression in some such terms as the following: given a probability α (say 0·01), the normal universe W_0 of the field and any other distribution W_i of the field,

a number n_1 can be found so that for $n \geqslant n_1$ the mean value of the test function for samples of n from W_1 will lie at or beyond the α probability point of the test function for samples of n from W_0: the smaller n_1 the more sensitive the test.

We consider, then, the infinite field of alternative tests of kurtosis represented by (3·1) when c assumes all positive values, and the infinite field of alternative universes represented by the Gram-Charlier frequency

$$\frac{1}{\sqrt{(2\pi)}} \exp\left\{\sum_{i=3}^{\infty} \frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\} e^{-\frac{1}{2}x^2}. \tag{5·1}$$

The universal variance is assumed to be unity, without loss of generality. The normal universe is a member of the field: it is found when all the λ_i $(i > 2)$ are zero. We assume that the conditions of § 2 are satisfied so that for indefinitely large samples the frequency distribution of $a(c)$ for all parent universes is normal. Obviously the efficiency of any particular test (i.e. $a(c)$ for a particular value of c) in regard to the normal and a particular non-normal alternative (i.e. a Gram-Charlier frequency with particular values of the λ_i) will be adjudged by considering the ratio of

(i) the difference between the universal mean values of $a(c)$ for the normal and the particular non-normal parent universes; to

(ii) the standard deviation of $a(c)$ for indefinitely large normal samples.

The most efficient test will be $a(c)$ for c a theoretically ascertainable function of the given λ_i which makes the ratio a maximum.

For indefinitely large samples the mean value ϕ of $a(c)$ when the parent universe is given by (5·1) is

$$\phi = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} dx\, |x|^c \exp\left\{\sum_i \frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\} e^{-\frac{1}{2}x^2}. \tag{5·2}$$

Obviously
$$\int_{-\infty}^{\infty} dx\, |x|^c \left(-\frac{d}{dx}\right)^{2m+1} e^{-\frac{1}{2}x^2} = 0.$$

Also, when $m \geqslant 1$,

$$\int_{-\infty}^{\infty} dx\, |x|^c \left(\frac{d}{dx}\right)^{2m} e^{-\frac{1}{2}x^2} = \left(\frac{c-1}{2}\right)! \, 2^{\frac{1}{2}(c+1)} c(c-2)(c-4)\ldots(c-2m+2), \tag{5·3}$$

a result readily inferable from the obvious fact that the left side vanishes for $c = 0, 2, \ldots, 2m-2$. Accordingly

$$\phi = \left(\frac{c-1}{2}\right)! \frac{2^{\frac{1}{2}(c+1)}}{\sqrt{(2\pi)}} \left\{1 + \frac{\lambda_4}{24} c(c-2) + \left(\frac{\lambda_3^2}{72} + \frac{\lambda_6}{720}\right) c(c-2)(c-4) + \ldots\right\}. \tag{5·4}$$

The normal value is given by the first term.

From (4·3), (4·5) and (4·6) it is evident that the value of the standard deviation, for larger normal samples (retaining only n^{-1}) is

$$\sigma = \frac{2^{\frac{1}{2}c}}{\sqrt{(\pi n)}} \left(\left(\frac{2c-1}{2}\right)! \sqrt{\pi} - \left(\frac{c-1}{2}!\right)^2 \frac{c^2+2}{2}\right)^{\frac{1}{2}}. \tag{5·5}$$

The principal term in the deviation $\phi - \phi^0$ (where ϕ^0 is the normal value), from (5·4), is

$$\delta = \frac{\frac{1}{2}(c-1)!\, 2^{\frac{1}{2}c}}{\sqrt{\pi}} \cdot \frac{\lambda_4 c(c-2)}{24}. \tag{5·6}$$

To a constant factor, the ratio δ/σ is given by the *first discriminant*

$$\rho(c) = c(c-2)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\frac{c-1}{2}!\right)^2} - \frac{c^2+2}{2}\right\}^{-\frac{1}{2}} \tag{5·7}$$

It will now be shown that $\frac{d\rho(c)}{dc} = 0$ for $c = 4$.

The discriminant may be written in the form

$$\rho(c) = c(c-2)\left(\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right)^{-\frac{1}{2}}, \tag{5·8}$$

where
$$I_c = \int_0^{\frac{1}{2}\pi} \cos^c\theta\, d\theta, \tag{5·9}$$

and
$$\frac{\rho'(c)}{\rho(c)} = \frac{1}{c} + \frac{1}{c-2} - \frac{1}{2}\left\{2^c\left(\frac{I_{2c}\log 2}{I_c} + \frac{I'_{2c}}{I_c} - \frac{I_{2c}I'_c}{I_c^2}\right) - c\right\} \Big/ \left\{\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right\}. \tag{5·10}$$

From (5·9)
$$J_c = I'_c = \int_0^{\frac{1}{2}\pi} d\theta \log^c\theta \log\cos\theta. \tag{5·11}$$

From a fairly well-known property

$$J_0 = \int_0^{\frac{1}{2}\pi} d\theta \log\cos\theta = -\tfrac{1}{2}\pi\log 2. \tag{5·12}$$

In (5·10) we shall be concerned only with even positive integer values of c. We have at once

$$I_0 = \frac{\pi}{2}, \quad I_2 = \frac{\pi}{4}, \quad I_4 = \frac{3\pi}{16}, \quad I_6 = \frac{5\pi}{32}, \quad I_8 = \frac{35\pi}{256}. \tag{5·13}$$

From (5·11)
$$J_{2c} = \int_0^{\frac{1}{2}\pi} d\theta \cos^{2c}\theta \log\cos\theta = \int_0^{\frac{1}{2}\pi} d(\sin\theta)\cos^{2c-1}\theta \log\cos\theta,$$

which, by partial integration,

$$= \int_0^{\frac{1}{2}\pi} d\theta \sin\theta\left(\overline{2c-1}\sin\theta\cos^{2c-2}\theta\log\cos\theta + \frac{\cos^{2c-1}\theta\sin\theta}{\cos\theta}\right)$$

$$= (2c-1)(J_{2c-2} - J_{2c}) + I_{2c-2} - I_{2c}.$$

Hence
$$2cJ_{2c} = (2c-1)J_{2c-2} - I_{2c} + I_{2c-2}. \tag{5·14}$$

From (5·12), (5·13) and (5·14),

$$\left.\begin{aligned} J_0 &= -\tfrac{1}{2}\pi\log 2, & J_6 &= (-60\pi\log 2 + 37\pi)/384, \\ J_2 &= (-2\pi\log 2 + \pi)/8, & J_8 &= (-840\pi\log 2 + 533\pi)/6144. \\ J_4 &= (-12\pi\log 2 + 7\pi)/64, \end{aligned}\right\} \tag{5·15}$$

Noting that $I'_{2c} = 2J_{2c}$ and substituting in the right side of (5·10) the values of I and J given by (5·13) and (5·15), we find $\rho'(4) = 0$. Table 7 gives the values of the discriminant for certain values of c.

The discriminant accordingly assumes a maximum value for $c = 4$, a result so remarkable that one might be inclined to suspect that it is a consequence of the form which was assumed for the alternative to the normal curve, a form which, in placing such emphasis on λ_4,

high-lights, so to speak, b_2 ($= \lambda_4 + 3$ when $\lambda_2 = 1$ for indefinitely large samples) as a test of normality. From the algebraic point of view this is anything but obvious: the property emerges from quite a complicated piece of algebra. It may also be emphasized that the field of alternatives (5·1) is not arbitrary; it is a general form of frequency distribution when all the λ_i are finite. Admittedly the discriminant takes account only of the term in λ_4 in the expansion; but this is certainly the most significant term for a wide class of frequency distributions, namely, those of homogeneous symmetrical functions of samples of n as n tends towards infinity under very general conditions for the parent universe, provided that the resulting frequency distribution can be assumed to have its third moment zero; for then the only term in n^{-1} in the frequency distribution of the function will be the term in λ_4. The significance of the property demonstrated must not be overstressed since it is subject to many qualifications, but it gives strong grounds for holding that, for very large samples, b_2 is the most efficient test of normality of tests of type $a(c)$ in relation to a very extended class of alternative universes. At the same time Table 7 shows that there can be little difference in efficiency in the field $a(c)$ for c ranging from close to 2 to about 5. There is but little doubt, on this showing, that b_2 is more sensitive than $a(1)$, a conclusion suggested on the basis of certain experimental results by E. S. Pearson (1935) and examined from the viewpoint of power function theory in § 6.

Table 7

$0<c<2$	Discriminant $\rho(c)$	$2<c<\infty$	Discriminant $\rho(c)$
$+0$	$-2{\cdot}334$	$2+0$	4·460
0·1	$-2{\cdot}541$	2·1	4·508
0·2	$-2{\cdot}725$	2·5	4·666
0·5	$-3{\cdot}188$	3·0	4·801
0·7	$-3{\cdot}441$	3·9	4·898
1·0	$-3{\cdot}758$	4·0	4·900
1·1	$-3{\cdot}851$	4·1	4·898
1·5	$-4{\cdot}166$	5·0	4·818
1·9	$-4{\cdot}405$	6·0	4·602
$2-0$	$-4{\cdot}460$	7·0	4·288
		8·0	3·906

Adverting to (5·4) in conjunction with (5·5), it might be asked if, on the analogy of the maximal property just demonstrated for the first discriminant, the function

$$\rho_2(c) = c(c-2)(c-4)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\left(\frac{c-1}{2}\right)!\right)^2} - \frac{c^2+2}{2}\right\}^{-1}$$

has a turning point at $c = 6$. The answer is in the negative. The value of $\rho_2'(6)/\rho_2(6)$ is, in fact, 15/34. At the same time there must be a zero of $\rho_2'(c)$ very near $c = 6$ since

$$\rho_2(5{\cdot}9) = 8{\cdot}79, \quad \rho_2(6) = 9{\cdot}20, \quad \rho_2(6{\cdot}1) = 8{\cdot}56.$$

Analogous to the field on tests of kurtosis represented by (3·1) we may consider as a field of tests of asymmetry:

$$g(c) = \frac{1}{n}\{-\Sigma' |x_i - \bar{x}|^c + \Sigma''(x_i - \bar{x})^c\} \Big/ \left\{\frac{1}{n}\Sigma(x_i - \bar{x})^2\right\}^{\frac{1}{2}c}, \tag{5·16}$$

where Σ' extends to the observations x_i less than the mean $\bar{x}$ and Σ'' to the rest of the sample. For $c = 3$ the test is, of course, $\sqrt{b_1}$. For normal samples

$$E\{g(c)\}^k = E\left\{-\frac{1}{n}\Sigma'|x_i-\bar{x}|^c+\frac{1}{n}\Sigma''(x_i-\bar{x})^c\right\}^k \Big/ E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}kc}, \tag{5·17}$$

the denominator of which is identical with the denominator of (4·1). Knowing the joint distribution (for normal samples) of $(x_1-\bar{x})$, $(x_2-\bar{x})$, ... (Geary, 1936), there is no theoretical difficulty in finding the mean values of the terms of the numerator for positive integer values of k. Here we shall be concerned only with the first and second moments, i.e. those for (5·17) for $k = 1$ and $k = 2$. We require the normal distribution of $z_1 = x_1-\bar{x}$ and the joint distribution of z_1 and $z_2 = x_2-\bar{x}$. These are

$$(z_1): \quad \left(\frac{n}{2\pi n-1}\right)^{\frac{1}{2}} \exp\left\{-\frac{nz_1^2}{2(n-1)}\right\} dz_1,$$

$$(z_1,z_2): \quad \frac{1}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}} \exp\left\{-\frac{(n-1)(z_1^2+z_2^2)}{2(n-2)}-\frac{z_1z_2}{(n-2)}\right\} dz_1dz_2 = f(z_1,z_2)\,dz_1dz_2. \tag{5·18}$$

Clearly the odd normal moments of $g(c)$ are zero. Then

$$E\left\{-\frac{1}{n}\Sigma'|x_i-\bar{x}|^c+\frac{1}{n}\Sigma''(x_i-\bar{x})^c\right\}^2 = \frac{n}{n^2}E|z_1|^{2c}+\frac{n(n-1)}{n^2}E_1(z_1,z_2), \tag{5·19}$$

where $E_1(z_1,z_2)$ is the mean value of the two-dimensional terms. We then have

$$\begin{aligned}E_1(z_1,z_2) &= \int_{-\infty}^{0}(-z_1)^c\,dz_1\int_{-\infty}^{0}(-z_2)^c\,dz_2f(z_1,z_2)-\int_{-\infty}^{0}dz_1(-z_1)^c\int_{0}^{\infty}dz_2z_2^cf(z_1,z_2)\\ &\quad-\int_{0}^{\infty}dz_1z_1^c\int_{-\infty}^{0}dz_2(-z_2)^cf(z_1,z_2)+\int_{0}^{\infty}dz_1z_1^c\int_{0}^{\infty}dz_2z_2^cf(z_1,z_2)\\ &= \int_0^\infty\int_0^\infty z_1^cz_2^c\,dz_1dz_2\{f(-z_1,-z_2)-f(-z_1,z_2)-f(z_1,-z_2)+f(z_1,z_2)\}\\ &= -\frac{2^{c+2}}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}}\frac{(n-2)^{c+1}}{(n-1)^{c+2}}\left(\frac{c}{2}!\right)^2\left\{1+\frac{(c+2)^2}{3!(n-1)^2}+\frac{(c+2)^2(c+4)^2}{5!(n-1)^4}+\ldots\right\},\end{aligned} \tag{5·20}$$

$$Ez_1^{2c} = \frac{2c-1}{2}!\left(\frac{2n-1}{n}\right)^c\frac{1}{\sqrt{\pi}}. \tag{5·21}$$

Also

$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^c = \left(\frac{2}{n}\right)^c\left(\frac{n+2c-3}{2}\right)!\Big/\left(\frac{n-3}{2}\right)!. \tag{5·22}$$

We now have all the expressions required for the variance of normal $g(c)$. We require, for what follows, only the term in n^{-1} which is

$$\sigma^2 = \frac{1}{n}\left\{\left(\frac{2c-1}{2}\right)!\frac{2^c}{\sqrt{\pi}}-\left(\frac{c}{2}!\right)^2\frac{2^{c+1}}{\pi}\right\}. \tag{5·23}$$

Consider now a field of alternative universes represented by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_3}{6}(x^3-3x)\right\}e^{-\frac{1}{2}x^2}, \tag{5·24}$$

the 'first approximation to the law of error' (for universal variance unity), obviously the most appropriate asymmetrical field, for different values of the parameter λ_3, and con-

taining as a member of the field the normal distribution found for $\lambda_3 = 0$. For indefinitely large samples from (5·24) the mean value of $g(c)$ is

$$\delta = \frac{2\lambda_3}{6\sqrt{(2\pi)}}\int_0^\infty dx\, x^c(x^3-3x)e^{-\frac{1}{2}x^2} = \frac{c}{2}!\,(c-1)\,2^{\frac{1}{2}c}\frac{\lambda_3}{3}\sqrt{\frac{1}{2\pi}}. \tag{5·25}$$

From (5·23) and (5·25)

$$\frac{\delta}{\sigma} = \frac{\lambda_3 n^{\frac{1}{2}}}{6}\tau(c), \tag{5·26}$$

the *skew discriminant* $\tau(c)$ being given by

$$\tau(c) = (c-1)\left\{\left(\frac{2c-1}{2}\right)!\left(\frac{c}{2}!\right)^{-2}\frac{\sqrt{\pi}}{2}-1\right\}^{-\frac{1}{2}} = (c-1)\left(\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right)^{-\frac{1}{2}}. \tag{5·27}$$

Log-differentiating,

$$\frac{\tau'(c)}{\tau(c)} = \frac{1}{c-1} - \frac{2^{c+1}}{2}\left\{\frac{1}{2c+1}\left(\frac{2J_{2c+2}}{I_{c+1}} - \frac{I_{2c+2}J_{c+1}}{I_{c+1}^2}\right)\right.$$
$$\left. - \frac{I_{2c+2}}{I_{c+1}}\frac{2}{(2c+1)^2} + \frac{I_{2c+2}}{I_{c+1}}\frac{\log 2}{2c+1}\right\}\left\{\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right\}^{-1}. \tag{5·28}$$

Setting $c = 3$ and using (5·13) and (5·15), we find that $\tau'(3) = 0$. Values of $\tau(c)$ for four values of c are as follows:

c	$\tau(c)$	c	$\tau(c)$
2	2·370	4	2·389
3	2·450	5	2·236

Accordingly, for indefinitely large samples the test of asymmetry $g(c)$ is most efficient for $c = 3$, when the test becomes the familiar $\sqrt{b_1}$. The margin in favour of this value of c, as compared with others in the range $2 \leqslant c \leqslant 5$, is, however, quite small.

6. TESTS OF KURTOSIS FROM THE POWER FUNCTION VIEWPOINT

It may be useful to open this section with an interpretation of the results of the previous section from the point of view of the type of error theory of J. Neyman & E. S. Pearson (1933, 1936). For this we consider two universes of the field, the normal W_0 and any non-normal universe W_1, and two tests of kurtosis $a(4) = b_2$ and $a(c_1)$ for a particular value c_1 of c. Suppose that samples are sufficiently large that $a(c)$, for samples from all universes of the field, may be regarded as normally distributed.

Given a probability α, a sample number n can be found so that the mean value of $a(c_1)$ from W_1 lies exactly at, say, the upper α probability point of the distribution of $a(c_1)$ from W_0. Then from the results established in the preceding section the value of $a(4)$ for the same sample of n from W_1 could lie beyond the α probability point of $a(4)$ for normal samples of n. Suppose that the rule adopted was to regard as non-normal all samples for which $a(c)$ lies beyond the normal α probability point, and suppose that a very large number N of samples were drawn, N_0 from universes not significantly different from normal (defining 'insignificance' in some manner) and N_1 from non-normal universes, so that $N = N_0 + N_1$, where N_0 and N_1 are not necessarily known in advance. Then using $a(c_1)$ the number of erroneous allocations will be approximately $\alpha N_0 + \frac{1}{2}N_1$, whereas using $a(4)$ the number will be $\alpha N_0 + (\frac{1}{2}-p)N_1$ $(\frac{1}{2} > p > 0)$, showing a definite advantage in favour of $a(4)$. The same conclusion emerges whatever value of $c \neq 4$ or whatever non-normal universe be taken for comparison.

The type of error approach reveals the theoretical weakness of using the method of § 5 for the assessment of relative efficiency of tests of normality; namely that the proportion of

errors of judgment, even using $a(4)$, remains large, due fundamentally to concentrating on a single value (the mean) as typical or representative of samples from the non-normal universe; it is also a disadvantage that the sample number n_1 is necessarily a function of the particular value c_1 of c. The method has further disadvantages of which the principal are perhaps (i) a somewhat restricted field of alternative universes; (ii) the assumption that the samples were indefinitely large, essential to justify the normality of $a(c)$ for samples from any member of the universe field.

The Neyman-Pearson power function approach which will now be considered cannot be regarded as entirely free from these objections in its application to the material so far available from this research. It enables us, at any rate, to contemplate samples which, if not small, are within the range of experimental practicability.

The problem of the relative efficiency of the different members of a field of tests of kurtosis $a(c)$ will now be considered in its power function aspects. For the present purpose the *power* may be defined as follows:

Given a probability α (say 0·01), a sample number n, a particular value c_1 of c and a non-normal parent universe W_1, the power, in relation to these data, represents the frequency of $a(c_1)$ for samples drawn at random from W_1 lying beyond the α probability point for $a(c_1)$ computed from samples drawn from a normal universe. The greater the power the more discriminatory the test. Accordingly, it is in theory necessary to know the frequency distribution of $a(c)$ for all sample sizes, for all values of c and for all universes. Considering that the only frequency distribution of the field contemplated which can be regarded as determined for all sample sizes is $a(1)$ for normal samples (Geary, 1935, 1936), many compromises are necessary to give any kind of practical effect to the power concept. The compromises proposed are as follows:

(1) The form $a_1(c)$, given by (3·2), is used instead of the form $a(c)$ given by (3·1).

(2) Only large samples are dealt with.

(3) The field of alternative universes is restricted.

Using $a_1(c)$, the first four moments (from the origin) of $a_1(c)$ for samples from any universe can be expanded without real difficulty, and so approximate frequency distributions (using the Karl Pearson or Gram-Charlier systems) can be obtained. As to (1), from experiments in $a(1)$ and $a(4)$ the writer has verified that, for medium-sized normal samples, there is little difference between the probability points (e.g. 0·01, 0·05) of $a_1(c)$ and $a(c)$, though the higher semi-invariants (given n) are larger for the latter. In regard to (2) and (3) little confidence could be reposed in the values of the moments computed from expansions even to n^{-3} unless the sample number was at least of the order of 100 when c is greater than, say, 3; and, even if the moments were known exactly, the empirical frequencies would be more than doubtful for small samples. The approach finds its main justification in the consideration that any errors due to these necessary compromises may be presumed to apply more or less equally and in the same direction to the tests of kurtosis compared; generous, perhaps too generous, advantage is taken of this justification in the concluding part of this section.

Set, then,
$$a_1(c) = \left\{\frac{1}{n}\Sigma\,|\,x_i\,|^c\right\} \Big/ \left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c}, \tag{6·1}$$

so that
$$\frac{a_1(c)}{\alpha} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}c}, \tag{6·2}$$

where
$$\alpha = \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad y_i = (|\,x_i\,|^c - \mu_{|c|})/\mu_{|c|}, \quad z_i = (x_i^2 - \mu_2)/\mu_2, \tag{6·3}$$

the universal mean being taken as zero, without loss of generality. Raising (6·2) to powers 1, 2, 3, 4, expanding to the required degree the final factor, multiplying by the first factor on the right, and setting down the mean value of each term we find, to n^{-3},

$$\begin{aligned}M_1'/\alpha = 1 &- \frac{1}{n}\{k_1^{(1)}(11) - k_2^{(1)}(02)\} + \frac{1}{n^2}\{k_2^{(1)}(12) - k_3^{(1)}[(03) + 3(11)(02)] + 3k_4^{(1)}(02)^2\} \\ &+ \frac{1}{n^3}\{k_3^{(1)}[3(11)(02) - (13)] + k_4^{(1)}[(04) - 3(02)^2 + 4(11)(03) + 6(12)(02)] \\ &- k_5^{(1)}[10(03)(02) + 15(11)(02)^2] + 15k_6^{(1)}(02)\}, \qquad (6{\cdot}4)\end{aligned}$$

$$\begin{aligned}M_2'/\alpha^2 = 1 &+ \frac{1}{n}\{k_2^{(2)}(02) - 2k_1^{(2)}(11) + (20)\} + \frac{1}{n^2}\{-k_3^{(2)}(03) + 3k_4^{(2)}(02)^2 \\ &+ 2k_2^{(2)}(12) - 6k_3^{(2)}(11)(02) - k_1^{(2)}(21) + k_2^{(2)}(20)(02) + 2k_2^{(2)}(11)^2\} \\ &+ \frac{1}{n^3}\{k_4^{(2)}[(04) - 3(02)^2] - 10k_5^{(2)}(03)(02) + 15k_6^{(2)}(02)^3 - 2k_3^{(2)}[(13) - 3(11)(02)] \\ &+ 4k_4^{(2)}[2(11)(03) + 3(12)(02)] - 30k_5^{(2)}(11)(02)^2 \\ &+ k_2^{(2)}[(22) - (20)(02)] - k_3^{(2)}[(20)(03) + 3(21)(02)] - 2k_2^{(2)}(11)^2 \\ &- 6k_3^{(2)}(12)(11) + 12k_4^{(2)}(11)^2(02) + 3k_4^{(2)}(20)(02)^2\}, \qquad (6{\cdot}5)\end{aligned}$$

$$\begin{aligned}M_3'/\alpha^3 = 1 &+ \frac{1}{n}\{k_2^{(3)}(02) - 3k_1^{(3)}(11) + 3(20)\} + \frac{1}{n^2}\{-k_3^{(3)}(03) + 3k_4^{(3)}(02)^2 \\ &+ 3k_2^{(3)}(12) - 9k_3^{(3)}(11)(02) - 3k_1^{(3)}(21) + 3k_2^{(3)}(20)(02) + 6k_2^{(3)}(11)^2 \\ &+ (30) - 3k_1^{(3)}(20)(11)\} + \frac{1}{n^3}\{k_4^{(3)}[(04) - 3(02)^2] - 10k_5^{(3)}(03)(02) + 15k_6^{(3)}(02)^3 \\ &- 3k_3^{(3)}[(13) - 3(11)(02)] + 6k_4^{(3)}[2(11)(03) + 3(12)(02)] - 45k_5^{(3)}(11)(02)^2 \\ &+ 3k_2^{(3)}[(22) - (20)(02)] - 3k_3^{(3)}[(20)(03) + 3(21)(02)] + 9k_4^{(3)}(20)(02)^2 \\ &- 6k_2^{(3)}(11)^2 - 18k_3^{(3)}(12)(11) + 36k_4^{(3)}(11)^2(02) - k_1^{(3)}(31) \\ &+ k_2^{(3)}(30)(02) + 3k_1^{(3)}(20)(11) \\ &+ 3k_2^{(3)}[(20)(12) + 2(21)(11)] - 9k_3^{(3)}(20)(11)(02) - 6k_3^{(3)}(11)^3\}, \qquad (6{\cdot}6)\end{aligned}$$

$$\begin{aligned}M_4'/\alpha^4 = 1 &+ \frac{1}{n}\{k_2^{(4)}(02) - 4k_1^{(4)}(11) + 6(20)\} + \frac{1}{n^2}\{-k_3^{(4)}(03) + 3k_4^{(4)}(02)^2 \\ &+ 4k_2^{(4)}(12) - 12k_3^{(4)}(11)(02) - 6k_1^{(4)}(21) + 6k_2^{(4)}(20)(02) + 12k_2^{(4)}(11)^2 \\ &+ 4(30) - 12k_1^{(4)}(20)(11) + 3(20)^2\} + \frac{1}{n^3}\{k_4^{(4)}[(04) - 3(02)^2] \\ &- 10k_5^{(4)}(03)(02) - 15k_6^{(4)}(02)^3 - 4k_3^{(4)}[(13) - 3(11)(02)] \\ &+ 8k_4^{(4)}[2(11)(03) + 3(12)(02)] - 60k_5^{(4)}(11)(02)^2 + 6k_2^{(4)}[(22) - (20)(02)] \\ &- 6k_3^{(4)}[(20)(03) + 3(21)(02)] + 18k_4^{(4)}(20)(02)^2 - 12k_2^{(4)}(11)^2 \\ &- 36k_3^{(4)}(12)(11) + 72k_4^{(4)}(11)^2(02) - 4k_1^{(4)}(31) + 4k_2^{(4)}(30)(02) \\ &+ 12k_1^{(4)}(20)(11) + 12k_2^{(4)}[(12)(20) + 2(21)(11)] \\ &- 36k_3^{(4)}(20)(11)(02) - 24k_3^{(4)}(11)^3 + (40) - 4k_1^{(4)}(30)(11) - 3(20)^2 \\ &- 6k_1^{(4)}(20)(21) + 3k_2^{(4)}(20)^2(02) + 12k_2^{(4)}(20)(11)^2\}, \qquad (6{\cdot}7)\end{aligned}$$

where $$k_r^{(p)} = \frac{\frac{1}{2}pc(\frac{1}{2}pc+1)(\frac{1}{2}pc+2)\ldots(\frac{1}{2}pc+r-1)}{r!}, \quad (fg) = Ey_i^f z_i^g,$$

the latter, of course, the same for all i. The (fg) required for the computation of (6·4)–(6·7) are

$$\left.\begin{aligned}
(11) &= (\mu_{|2+c|}-\mu_2\mu_{|c|})/\mu_2\mu_{|c|},\\
(02) &= (\mu_4-\mu_2^2)/\mu_2^2,\\
(12) &= (\mu_{|4+c|}-2\mu_{|2+c|}\mu_2-\mu_{|c|}\mu_4+2\mu_{|c|}\mu_2^2)/\mu_{|c|}\mu_2^2,\\
(03) &= (\mu_6-3\mu_4\mu_2+2\mu_2^3)/\mu_2^3,\\
(04) &= (\mu_8-4\mu_6\mu_2+6\mu_4\mu_2^2-3\mu_2^4)/\mu_2^4,\\
(13) &= [\mu_{|6+c|}-3\mu_{|4+c|}\mu_2+3\mu_{|2+c|}\mu_2^2-\mu_{|c|}(\mu_6-3\mu_4\mu_2+3\mu_2^3)]/\mu_{|c|}\mu_2^3,\\
(21) &= [\mu_{|2c+2|}-2\mu_{|c+2|}\mu_{|c|}-\mu_2(\mu_{|2c|}-2\mu_{|c|}^2)]/\mu_{|c|}^2\mu_2,\\
(22) &= (\mu_{|2c+4|}-2\mu_{|2c+2|}\mu_2+\mu_{|2c|}\mu_2^2-2\mu_{|c+4|}\mu_{|c|}+4\mu_{|c+2|}\mu_{|c|}\mu_2-3\mu_{|c|}^2\mu_2^2+\mu_{|c|}^2\mu_4)/\mu_{|c|}^2\mu_2^2,\\
(20) &= (\mu_{|2c|}-\mu_{|c|}^2)/\mu_{|c|}^2,\\
(30) &= (\mu_{|3c|}-3\mu_{|2c|}\mu_{|c|}+2\mu_{|c|}^3)/\mu_{|c|}^3,\\
(31) &= (\mu_{|3c+2|}-3\mu_{|2c+2|}\mu_{|c|}+3\mu_{|c+2|}\mu_{|c|}^2-\mu_{|3c|}\mu_2+3\mu_{|2c|}\mu_{|c|}\mu_2-3\mu_{|c|}^3\mu_2)/\mu_{|c|}^3\mu_2,\\
(40) &= (\mu_{|4c|}-4\mu_{|3c|}\mu_{|c|}+6\mu_{|2c|}\mu_{|c|}^2-3\mu_{|c|}^4)/\mu_{|c|}^4.
\end{aligned}\right\} \quad (6\cdot8)$$

(6·8) is, of course, an immediate consequence of (6·3). The writer has checked the accuracy of formulae (6·4)–(6·7) by reference to the normal universe for $c = 1$.

The reader will have no illusions as to the magnitude of the task of applying the foregoing theory to particular cases. The formulae are set down, however, in the hope that other researchers will be sufficiently sensible of the importance of the theory to assist in building up a fairly extensive set of results. The writer has to be content, in the meantime, to consider the case of the symmetrical universe field given by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_4}{24}\left(\frac{d}{dx}\right)^4\right\}e^{-\frac{1}{2}x^2}, \quad (6\cdot9)$$

when $\lambda_4 = \frac{1}{2}$, the normal being given, of course, for $\lambda_4 = 0$, and for $c = 4$ and $c = 1$. These values of c are selected because the theory in § 5 has suggested that $a(4)$ is probably the most efficient of the test-field $a(c)$, while $a(1)$ is the only member of the field for which the normal

Table 8. *Moments from formulae* (6·8)

(fg)	$c=4$		$c=1$	
	Normal	$\lambda_4=\frac{1}{2}$	Normal	$\lambda_4=\frac{1}{2}$
(11)	4	5·428571	1	1·17021276
(02)	2	2·5	2	2·5
(12)	24	45·64286	3	4·88297871
(03)	8	14	8	14
(04)	60	138	60	138
(13)	216	544·2857	21	44·106383
(21)	256/3	177·71428	1·141593	1·75544898
(22)	2,720/3	2,481·92857	7·707963	14·766814
(20)	32/3	16·142857	0·570796	0·63834981
(30)	352	799·142857	0·429204	0·6405182
(31)	4,352	12,785·2653	3	5·236134
(40)	23,552	73,250·178	—	2·002492

distribution is known for samples of all sizes. The necessary moments (fg) given by (6·8) are shown in Table 8. Based on the values in this table, moments (M') given by (6·4)–(6·7) of $a_1(c)$ and semi-invariants (L) derived therefrom are as follows. The normal values are, of course, known exactly but were computed for the purpose of checking the formulae:

$c = 4$; *normal universe*

$$\frac{L_1}{3} = \frac{M'_1}{3} \simeq 1 - \frac{2}{n} + \frac{4}{n^2} - \frac{8}{n^3},$$

$$\frac{M'_2}{9} \simeq 1 - \frac{4}{3n} - \frac{28}{n^2} + \frac{1040}{3n^3}, \quad \frac{L_2}{9} \simeq \frac{8}{3n} - \frac{40}{n^2} + \frac{1136}{3n^3},$$

$$\frac{M'_3}{27} \simeq 1 + \frac{2}{n} - \frac{48}{n^2} - \frac{1040}{n^3}, \quad \frac{L_3}{27} \simeq \frac{64}{n^2} - \frac{2368}{n^3},$$

$$\frac{M'_4}{81} \simeq 1 + \frac{8}{n} + \frac{40}{3n^2} - \frac{3520}{n^3}, \quad \frac{L_4}{81} \simeq \frac{3840}{n^3}.$$

$c = 4$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{3 \cdot 5} = \frac{M'_1}{3 \cdot 5} \simeq 1 - \frac{3 \cdot 357}{n} + \frac{11 \cdot 822}{n^2} + \frac{12 \cdot 1}{n^3},$$

$$\frac{M'_2}{(3 \cdot 5)^2} \simeq 1 - \frac{2 \cdot 286}{n} - \frac{57 \cdot 34}{n^2} + \frac{776 \cdot 03}{n^3}, \quad \frac{L_2}{(3 \cdot 5)^2} \simeq \frac{4 \cdot 4286}{n} - \frac{92 \cdot 25}{n^2} + \frac{831 \cdot 2}{n^3},$$

$$\frac{M'_3}{(3 \cdot 5)^3} \simeq 1 + \frac{3 \cdot 215}{n} - \frac{107 \cdot 47}{n^2} - \frac{2853 \cdot 89}{n^3}, \quad \frac{L_3}{(3 \cdot 5)^3} \simeq \frac{144 \cdot 61}{n^2} - \frac{6193 \cdot 95}{n^3},$$

$$\frac{M'_4}{(3 \cdot 5)^4} \simeq 1 + \frac{13 \cdot 143}{n} + \frac{20 \cdot 49}{n^2} - \frac{9529}{n^3}, \quad \frac{L_4}{(3 \cdot 5)^4} \simeq \frac{10{,}587}{n^3}.$$

$c = 1$; *normal universe*

$$L_1 = M'_1 \simeq 0 \cdot 7978845608 + \frac{0 \cdot 19947114}{n} + \frac{0 \cdot 02493389}{n^2} - \frac{0 \cdot 03116737}{n^3},$$

$$L_2 \simeq \frac{0 \cdot 04507034}{n} - \frac{0 \cdot 07957747}{n^2} + \frac{0 \cdot 03978874}{n^3},$$

$$L_3 \simeq -\frac{0 \cdot 01685645}{n^2} + \frac{0 \cdot 07613597}{n^3}.$$

$c = 1$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{\mu_{|1|}} = \frac{M'_1}{\mu_{|1|}} \simeq 1 + \frac{0 \cdot 35239362}{n} - \frac{0 \cdot 159616}{n^2} - \frac{0 \cdot 745838}{n^3},$$

$$\frac{M'_2}{\mu^2_{|1|}} \simeq 1 + \frac{0 \cdot 79792429}{n} - \frac{0 \cdot 458012}{n^2} - \frac{1 \cdot 800648}{n^3}, \quad \frac{L_2}{\mu^2_{|1|}} \simeq \frac{0 \cdot 09313705}{n} - \frac{0 \cdot 262961}{n^2} - \frac{0 \cdot 196477}{n^3},$$

$$\frac{M'_3}{\mu^3_{|1|}} \simeq 1 + \frac{1 \cdot 336592}{n} - \frac{0 \cdot 850081}{n^2} - \frac{3 \cdot 239101}{n^3}, \quad \frac{L_3}{\mu^3_{|1|}} \simeq \frac{0 \cdot 053356}{n^2} + \frac{0 \cdot 204164}{n^3},$$

$$\mu_{|1|} = 0 \cdot 78126197.$$

Two sample sizes were considered: $n = 100$ and $n = 500$. For $n = 100$ and $c = 4$, the

following are the Pearson Type IV frequencies of $a_1(4)$ when the parent universes are normal and have $\lambda_4 = \beta_2 - 3 = \frac{1}{2}$ respectively:

$$\text{Normal: } \lambda_4 = 0. \quad \left.\begin{aligned} &\kappa \cos^{11\cdot3350}\theta\, e^{13\cdot015436}\, dx, \\ &\tan\theta = (x - 1\cdot873387)/0\cdot765849, \\ &\log_{10}\kappa = \bar{3}\cdot2644596. \end{aligned}\right\} \tag{6·10}$$

$$\lambda_4 = \tfrac{1}{2}: \quad \left.\begin{aligned} &\kappa \cos^{6\cdot0096}\theta\, e^{2\cdot3128\theta}\, dx, \\ &\tan\theta = (x - 2\cdot8522)/0\cdot9062, \\ &\log_{10}\kappa = \bar{1}\cdot7499974. \end{aligned}\right\} \tag{6·11}$$

The normal probability points shown in column (2) of Table 10 were derived from the foregoing normal frequency (6·10); the points in column (3) were derived from a Gram-Charlier formula (Geary, 1935). The 0·01 and 0·05 points given in column (2) are practically identical with those given by E. S. Pearson (1929) for $a(4)$, namely, 4·39 and 3·77. The powers given in column (4) are the aggregate frequencies lying beyond the values of the variate shown in column (2) on the assumption that the actual frequency was (6·11). The corresponding figures for $c = 1$ given in column (5) were based on a Gram-Charlier formula.

Table 9. *Power of $a_1(c)$ for $c = 4$ and $c = 1$ of discriminating (6·9) for $\lambda_4 = \frac{1}{2}$ from the normal ($\lambda_4 = 0$) at four normal theory probability levels. Samples of* 100

Normal theory probability (1)	Normal theory probability points		Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$	
	$c=4$ (upper) (2)	$c=1$ (lower) (3)	$c=4$ (4)	$c=1$ (5)
0·01	4·3836	0·7482	0·0648	0·0695
0·05	3·7744	0·7642	0·1995	0·1979
0·10	3·5195	0·7725	0·3163	0·3037
0·20	3·3110	0·7824	0·4525	0·4597

Before discussing the comparative powers in Table 9 it will be convenient to give a table, 11, on the same lines but for $n = 500$. On account of the larger sample size it has been necessary to change the reference-probabilities given in column (1). For the construction of this table Gram-Charlier formulae were used throughout—the probability points being determined from the E. A. Cornish & R. A. Fisher (1937) formulae—after verifying that for two of the probability levels, 0·01 and 0·05, the probability points for $c = 4$ (column (2) above) did not differ appreciably from those given by E. S. Pearson, namely, 3·60 and 3·37 (for $a(4)$), based on a Type IV curve.

The analysis in § 5 has enabled us to come fairly firmly to the conclusion that for indefinitely large samples $a(4)$ was to be preferred to $a(1)$ as a test of normality. We see from Tables 9 and 10 that this is subject to an important qualification. Table 9 shows that the discriminating power is definitely greater for samples of 500 for $a(4)$ than for $a(1)$, but the superiority is less emphatic than might have been anticipated from § 5. For medium-sized samples (Table 9) $a(4)$ exhibits no superiority. Of course, these conclusions are very tentative, as being based upon a single alternative and on particular sample sizes. The writer had proposed, in addition, to examine the universes (i) $\lambda_3 = 0$, $\lambda_4 = 1$ and (ii) $\lambda_3^2 = \lambda_4 = \frac{1}{2}$ as alternatives to the normal but time did not permit; he ventures to repeat the hope that other students will take the matter up.

Table 10. *Power of $a_1(c)$ for $c = 4$ and $c = 1$ of discriminating (6·9) for $\lambda_4 = \frac{1}{2}$ from the normal ($\lambda_4 = 0$) at four probability levels. Samples of 500*

Normal probability (1)	Normal probability points		Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$	
	$c=4$ (upper) (2)	$c=1$ (lower) (3)	$c=4$ (4)	$c=1$ (5)
0·005	3·7062	0·773167	0·1934	0·2067
0·01	3·6094	0·775684	0·2920	0·2790
0·05	3·3766	0·782482	0·5955	0·5196
0·10	3·2695	0·786058	0·7392	0·6509

7. CONCLUSION AND SUMMARY

In § 2 of the present paper it is shown that the actual probability of differences between means and variances derived from random samples on the nul-hypothesis may differ considerably from the probability derived from the standard tables (compiled on the assumption that the universal distribution is normal), when, in fact, the universal distribution is *not* normal. Accordingly, the standard tables cannot validly be used unless tests, based on the sample from which the inferences are to be drawn, or on a series of samples produced under similar conditions, have established the likelihood that the universal distribution is approximately normal. In certain cases—but these must be few—the nature of the material may, of itself, suffice to justify the assumption of universal normality. When universal normality cannot be assumed, the best course will be to correct the standard tables using, for this purpose, the moments (up to, say, the fourth) derived from the sample, in conjunction with the formulae given in § 2. This procedure is, of course, open to the objection that the moments derived from the sample may, in fact, differ substantially from the (in general unknown) universal moments, so that any probabilistic inference derived using sample moments must be accepted with reserve. If $b_2 = 3{\cdot}5$, say, it would be safer to assume that the universal value β_2 is 3·5, than to hope (without other evidence) that it is 3, the normal value; it might be 3·75 or even 4, when, usually, the standard table probabilities will be still further astray. It should not be difficult to construct supplementary tables giving very approximate corrections of the standard tables, using the moment expansions given in § 2, for different values of $\sqrt{\beta_1}$ and β_2. To compute unbiassed estimates of the latter, R. A. Fisher's k statistics (1929) should, of course, be used.

It may be asked if testing for normality and, when necessary, correction for universal non-normality is worth the trouble. To answer this question it is desirable to have regard to the logical position of the statistician, concerned with drawing inferences from samples, whose characteristic approach may be defined as *reductio ad paene absurdum*: if an event is highly improbable it must be regarded for practical purposes as impossible. St Thomas Aquinas's* famous 'certitude of probability' is peculiarly apt as applied to the mental attitude of the statistician, from two quite different viewpoints. The first is that decision, and action based on that decision, for which there is not certainty, but merely probabilistic preference, is absolute. One does not say that one has a preference of 20 to 1 for Fertilizer A

* 'According to the Philosopher, certitude is not to be sought equally in every matter....Hence the certitude of probability suffices, such as may reach the truth in the greater number of cases, although it fails in the minority' (*Summa* 11a–11ae q. lxx, a. 2).

over Fertilizer B because the differences between the yields is at or near the 5 % probability point of some test functions: one necessarily decides without qualification that A is better than B.

The second aspect, which has the greater relevance in the present case, is that the statistician regards himself as endowed with 'certitude' when he knows that if he repeated an experiment, as to, say, significant differences in averages, a great number of times, he would be in error in attributing significant difference when, in fact, there was none, in a predetermined proportion of cases. He has certitude as to the probability though his decision in the individual case may be wrong. What is curious is that decisions (which, in effect, are absolute) can be based on probability levels which vary with the temperament of the statistician from perhaps a conservative 0·001 to a daring 0·1. For the particular statistician the probability level will vary with the case: for instance, the present writer would be inclined to suspect non-normality near the 10 % probability level of the $a(1)$ table, whereas he would not be disposed to attach significance in, say, analysis of variance, until about the $2\frac{1}{2}$ % level. Naturally the level will depend on the importance attaching to the decision.

Since all the statistician usually requires from the table of probability for a given measure of significance is whether, on the nul-hypothesis, the probability is 'small', absolute precision is not necessary in the probability. If the probability is thought to be minute, say 0·001, it does not matter if in actual fact it is 0·002 or 0·0005. If, on the contrary, the standard table value is approaching the statistician's level of decision it surely matters a great deal: if he thinks his judgment is likely to be erroneous in 1 out of 20 experiments it must be of importance if, in fact, the true probability is something like 1 in 10 or 1 in 5. These are the kinds of contrasts that appear from § 2, from comparison of standard table probabilities with 'actual' probabilities found when the samples were assumed to be randomly drawn from certain arbitrarily selected types of non-normal universes. The computed probabilities in § 2 admittedly make no claim to exactitude in most of the cases, since the formulae were strained by their application to small sample theory. The point is, however, that the estimates of the actual probabilities are unbiassed in regard to the 'normal theory' probabilities: if the former could be closer to the latter, they might also be further away.

There is one case which is in a quite exceptional category, namely that considered at the beginning of § 2. As far as the writer is aware, this case has never been examined *theoretically* before, despite the extreme simplicity of the algebra. It is shown that in the simplest case of analysis of variance, when the two sample numbers are of the same order of magnitude, the variance is proportional, approximately, to $(\beta_2 - 1)$, so that quite a small measure of universal kurtosis materially changes the probability. Statisticians must have been affected by a kind of hypnosis in favour of normal theory to have overlooked so trivial a point, a stricture from which the writer is not particularly concerned to exclude himself! An exception was E. S. Pearson (1931) who, on the basis of his results cited in § 2 (*a*), sounded a warning: 'The illustration should serve to emphasize the fact that certain of the "normal theory" tests can be used with greater confidence than others when dealing with samples from populations whose distribution laws are not known.'

An interesting chapter could be written on the fluctuations in the attitude of statisticians during the past century on the question of the occurrence of the normal frequency distribution in nature, a chapter, perhaps, in a large work on Fashions in the Sciences down the Ages. Amongst the following the historian may find the reasons for the prejudice in favour of the hypothesis of universal normality up to, say, the end of the last century:

(1) The fact that, to a close approximation, it applies in a wide range of *mathematical* conditions.

(2) The fact that the theory found practical applications predominantly in assessing the probability of errors in astronomical measurements and in games of chance where the mathematical model could reasonably be assumed to apply.

(3) The beauty of the mathematical theory and the facility of algebraic manipulation in the function involved.

(4) The general shape to the visual sense of such frequency distributions as were known, before χ^2 imposed its discipline.

With the development, about the beginning of the century, of the theory of moments, statisticians became almost over-conscious of universal non-normality. The concomitant semi-invariant approach had quite a different background. The difference between the moment and Karl Pearson curve system on the one hand and semi-invariants and the Gram-Charlier system on the other is fundamentally that for the former normality is a particular case like any other, whereas for the latter normality is basic and generative. Each system has its advantages and disadvantages as applied to the determination of frequency distributions of which the lower moments are known. In fanciful terms one might say that in the ship Gram-Charlier one might sail in perfect safety but only within limited, and more or less ascertainable, range of Port Normality, whereas in the good craft Pearson one can sail the seven seas—at one's own risk.*

Our historian will find a significant change of attitude about a quarter-century ago following on the brilliant work of R. A. Fisher who showed that, when universal normality could be assumed, inferences of the widest practical usefulness could be drawn from samples of any size. Prejudice in favour of normality returned in full force and interest in non-normality receded to the background (though one of the finest contributions to non-normal theory was made during the period by R. A. Fisher himself), and the importance of the underlying assumptions was almost forgotten. Even the few workers in the field (amongst them the present writer) seemed concerned to show that 'universal non-normality doesn't matter': we so wanted to find the theory as good as it was beautiful. References (when there were any at all) in the text-books to the basic assumptions were perfunctory in the extreme. Amends might be made in the interest of the new generation of students by printing in leaded type in future editions of existing text-books and in all new text-books:

Normality is a myth; there never was, and never will be, a normal distribution.

This is an over-statement from the practical point of view, but it represents a safer initial mental attitude than any in fashion during the past two decades.

As already indicated, the present work is incomplete, especially on the experimental side. The writer hopes that he has created a *prima facie* case for the importance of testing for normality.

SUMMARY

(i) Inferences drawn from the standard (normal) tables of z and t may be seriously in error if the conditions in which the standard tables apply (the principal of which is that the universes from which the samples are drawn are normal) are ignored.

* This comment must not be taken as applying to the problem of curve-fitting, i.e. to fitting a smooth curve to given frequencies, but to the problem of estimating the frequency function given the first few semi-invariants.

(ii) Sufficient conditions are given for the approach to normality, with increasing sample size, of the field of tests of normality $a(c)$ (given by (3·1)) for $c > 0$.

(iii) Many term expansions of the first four moments of $a(c)$ for normal samples are given with practical applications designed to find the values of c for which the moments could be used with confidence to find the frequency distributions for medium-size samples; semi-invariants of $a_1(2{\cdot}4)$ and $a_1(4)$ ($a_1(c)$ is given by (3·2)) are compared; correlations between $a_1(c)$ and $a_1(c')$ are examined.

(iv) For indefinitely large samples and a wide field of alternative universes $a(4)$ is found to be the most sensitive test of kurtosis and an analogous test of asymmetry $g(c)$ is found to be most sensitive for $c = 3$, $g(3)$ being the familiar $\sqrt{b_1}$.

(v) An examination of the relative efficiency of $a(1)$ and $a(4)$ from the Power Function point of view suggests that $a(4)$ is increasingly to be preferred as the sample size increases; for samples of moderate size $a(1)$ is probably as efficient as $a(4)$.

(vi) Throughout the paper a considerable range of formulae is given in case students may feel interested to carry the writer's researches a stage further so as to give a firmer basis to his conclusions or to modify them. It is suggested (§ 4) that the preparation of a table of probability points of $a(2{\cdot}2)$ for normal samples of different sizes be taken in hand.

REFERENCES

Baker, G. A. (1932). *Ann. Math. Statist.* **3**, 1.
Bartlett, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 226.
Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. Int. Statist.* **5**, 307.
Craig, C. C. (1928). *Metron*, **7**, 3.
Eden T. & Yates, F. (1933). *J. Agric. Sci.* **23**, 6.
Fisher, R. A. (1925). *Metron*, **5**, 90.
Fisher, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
Fisher, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
Fisher, R. A. & Wishart, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.
Fréchet, M. (1937). *Généralités sur les Probabilités. Variables aléatoires.*
Geary, R. C. (1933). *Biometrika*, **25**, 184.
Geary, R. C. (1935). *Biometrika*, **27**, 310, 353.
Geary, R. C. (1936). *J. Roy. Statist. Soc.* (*Supplement*), **3**, 178.
Geary, R. C. (1936). *Biometrika*, **28**, 295.
Geary, R. C. (1947). *Biometrika*, **34**, 68.
Geary, R. C. & Pearson, E. S. (1938). *Tests of Normality.*
Geary, R. C. & Worlledge, J. P. G. (1946). *Biometrika*, **34**, 98.
Gosset, W. S. (1908). *Biometrika*, **6**, 1.
Hsu, C. T. & Lawley, D. N. (1940). *Biometrika*, **31**, 238.
Kendall, M. G. (1941). *Biometrika*, **32**, 81.
Nair, A. N. K. (1942). *Sankhyā*, **5**, 393.
Neyman, J. & Pearson, E. S. (1933). *Philos. Trans.* A, **231**, 289.
Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 1.
Pearson, E. S. (1929). *Biometrika*, **21**, 337.
Pearson, E. S. (1930). *Biometrika*, **22**, 239.
Pearson, E. S. (1931*a*). *Biometrika*, **22**, 423.
Pearson, E. S. (1931*b*). *Biometrika*, **23**, 114.
Pearson, E. S. (1935). *Biometrika*, **27**, 333.
Pearson, E. S. & Adyanthaya, N. K. (1929). *Biometrika*, **21**, 259.
Pearson, Karl (1895). *Philos. Trans.* A, **186**, 343.
Pepper, Joseph (1932). *Biometrika*, **24**, 55.
Rider, P. R. (1931). *Ann. Math. Statist.* **2**, 48.
Rietz, H. L. (1939). *Ann. Math. Statist.* **10**, 265.
Shewhart, W. A. & Winters, F. W. (1928). *J. Amer. Statist. Ass.* **23**, 144.
Wishart, J. (1930). *Biometrika*, **22**, 224.
Yasukawa, K. (1934). *Tokohu Math. J.* **38**, 465.

THE STRATIFIED SEMI-STATIONARY POPULATION

By S. VAJDA

1. Constant population

Let a set of non-increasing real values $p_0 = 1,\ p_1, \ldots, p_n, p_{n+1} = 0$ be given, and let p_i represent the probability of a person of age 0 surviving the i following years. Further, let $l_0, l_1, \ldots, l_n$ represent the numbers of persons of age $0, 1, \ldots, n$ living at time $t = 0$. We consider then the development of such a population during the years following $t = 0$, under the assumption that the probabilities p_i remain the same throughout the period investigated.

Only persons of age 0 are to enter the population, and the number of such entrants shall be such that the total of the population is kept constant at a number $H = \sum_{i=0}^{n} l_i$. At the end of the first year the survivors of the H persons who were alive at $t = 0$ will be (if we put $l_i/p_i = r_i$, say)

$$\sum_{i=0}^{n-1} l_i \frac{p_{i+1}}{p_i} = \sum_{i=0}^{n-1} r_i p_{i+1} \leqslant H,$$

and therefore the number of entrants at the beginning of the second year (i.e. at $t = 1$) is $\phi_1 = H - \sum_{i=0}^{n-1} r_i p_{i+1}$.

By the same argument the entrants at $t = 2$ will be

$$\phi_2 = H - \phi_1 p_1 - \sum_{i=0}^{n-2} r_i p_{i+2},$$

and so on; generally

$$\phi_t = H - \phi_{t-1} p_1 - \phi_{t-2} p_2 - \ldots - \phi_1 p_{t-1} - \sum_{i=0}^{n-t} r_i p_{i+t}, \tag{1}$$

as long as $t \leqslant n$, that is, as long as there are survivors of the initial population. For $t > n$ we obtain

$$H = \phi_t + \phi_{t-1} p_1 + \phi_{t-2} p_2 + \ldots + \phi_{t-n} p_n. \tag{2}$$

We want to find an expression for ϕ_t, which must obviously depend on $l_1, l_2, \ldots, l_n$. Now (2) is a difference equation for the function ϕ_t of t and can easily be solved. For this purpose consider the 'characteristic equation'

$$x^n + x^{n-1} p_1 + x^{n-2} p_2 + \ldots + x p_{n-1} + p_n = 0. \tag{3}$$

Let this equation have the roots $x_1, x_2, \ldots, x_r$, where x_i is a k_i-fold root and $x_i \neq x_j$. We have then as a solution of the difference equation (2)

$$\phi_t = H_1 + P_1(t)\, x_1^t + \ldots + P_r(t)\, x_r^t, \tag{4}$$

where $H_1 = H/\Sigma p_i$ and $P_i(t) = \alpha_{i1} + \alpha_{i2} t + \ldots + \alpha_{ik_i} t^{k_i - 1}$. The α_{ij} must be found from the initial population, i.e. from equations of the form (1) which contain the first n numbers of entrants $\phi_1, \phi_2, \ldots, \phi_n$. But we find by inspecting these equations, which are of the form ($t \leqslant n$)

$$H = \phi_t + \phi_{t-1} p_1 + \ldots + \phi_1 p_{t-1} + r_0 p_t + r_1 p_{t+1} + \ldots + r_{n-t} p_n,$$

that they are equivalent to

$$r_t = \phi_{-t} = H_1 + \sum_{i=1}^{r} \sum_{j=1}^{k_i} \alpha_{ij} x_i^{-t} (-t)^{j-1}. \tag{5}$$

Hence the α_{ij} can be fixed, dependent on the $r_t = l_t/p_t$ and thus on the initial population.

We have thus proved:

If a population with an age distribution $l_0, l_1, \ldots, l_n$ is subject to survival rates p_i $(i = 1, 2, \ldots, n)$, and if this population is kept constant by ϕ_t entrants of age 0 at the end of the tth year, then ϕ_t is given by (4), *where the x_i are the different roots of* (3), *and the α_{ij} must be found from the set* (5).

The population after t years will have the following age distribution:

$$\phi_t, \quad \phi_{t-1}p_1, \quad \phi_{t-2}p_2, \quad \ldots, \quad \phi_{t-n}p_n.$$

It can easily be proved that, if the p_t are decreasing (and not merely non-increasing), then for all the roots x_i of equation (3) we have $|x_i| < 1$ and that any real root must be negative. Hence the ϕ_t will oscillate around their limit $\lim_{t=\infty} \phi_t = H_1$. The age distribution of the population thus tends, again through oscillations, to $H_1, H_1p_1, \ldots, H_1p_n$, which may be called the *intrinsic stationary population*. Obviously, if the initial population has already this distribution, it will not alter any more and the number of entrants will be constant and $= H_1$. In such a case all $\alpha_{ij} = 0$, and $r_t = H_1 = l_0$, whatever the x_i may be.

On the other hand, if $p_t = p_{t+1}$ holds for one or more values of i, then we may get cycles, and this is easily seen for the equation $x^n + x^{n-1} + \ldots + x + 1 = 0$. All roots have modulus 1, and it depends on the initial population whether we are dealing with the stationary case or with periodic cycles. No tendency towards an intrinsic stationary population appears in such a case.

Example. Let us assume that we have the following probabilities of survival:

p_1	p_2	p_3	p_4	p_5	p_6
7/8	49/96	5/32	13/384	1/384	0

The characteristic equation can then be written

$$384x^5 + 336x^4 + 196x^3 + 60x^2 + 13x + 1 = 0,$$

which has the five different roots

$$-\tfrac{1}{4} \pm \sqrt{-\tfrac{5}{48}}, \quad -\tfrac{1}{8} \pm \sqrt{-\tfrac{7}{64}} \quad \text{and} \quad -\tfrac{1}{8}.$$

The initial population will be assumed to be

l_0	l_1	l_2	l_3	l_4	l_5
859	1269	229	50	115	56

which implies $H_1 = 1000$ (approx.). Therefore the $r_t = l_t/p_t$ are

r_0	r_1	r_2	r_3	r_4	r_5
859	1450·38	448·86	319·14	3405·42	21420·90

From $r_k = 1000 + \alpha_1 x_1^{-k} + \alpha_2 x_2^{-k} + \ldots + \alpha_5 x_5^{-k}$ $(k = 1, 2, \ldots, 5)$ we find

α_1	α_2	α_3	α_4	α_5
$-70 + 60i$	$-70 - 60i$	0	0	-1

It follows that the number of entrants in the year t $(= 0, 1, 2, \ldots)$ will be

$$\phi_t = 1000\left[1 + (-0{\cdot}07 + 0{\cdot}06i)\left(-\tfrac{1}{4} + \sqrt{-\tfrac{5}{48}}\right)^t + (-0{\cdot}07 - 0{\cdot}06i)\left(-\tfrac{1}{4} - \sqrt{-\tfrac{5}{48}}\right) + \frac{-0{\cdot}001}{(-8)^t}\right].$$

These numbers are given in the first row of Table 1, which shows the evolution of the whole population.

Table 1

t	0	1	2	3	4	5	6	7	8 and after
Age 0	859	996	1025	989	1001	1001	999	1000	1000
1	1269	752	872	897	865	876	876	874	875 (= 1000 × 7/8)
2	229	740	438	508	523	505	511	511	510 (= 1000 × 49/96)
3	50	70	227	134	156	160	155	156	156 (= 1000 × 5/32)
4	115	11	15	49	29	34	34	34	34 (= 1000 × 13/384)
5	56	9	1	1	4	2	3	3	3 (= 1000 × 1/384)
	2578	2578	2578	2578	2578	2578	2578	2578	2578

2. TWO CONSTANT POPULATIONS

All this covers well known ground.* A new problem arises, however, when we consider two initial populations with two sets of probabilities of survival, say p_i $(i = 1, 2, \ldots, n_1)$ and $\overline{p}_i$ $(i = 1, 2, \ldots, n_2)$, where $p_0 = \overline{p}_0 = 1$ and $p_{n_1} \cdot p_{n_2} \neq 0$. We ask now whether it is possible to keep both constant by the same number of yearly entrants. More precisely:

Let the two equations

$$\sum_{i=0}^{n_1} p_i x^{n_1 - i} = 0 \quad \text{and} \quad \sum_{i=0}^{n_2} \overline{p}_i y^{n_2 - 1} = 0$$

have the roots $x_1, \ldots, x_r$ with multiplicities $k_1, \ldots, k_r$ and $y_1, \ldots, y_s$ with multiplicities $j_1, \ldots, j_s$ respectively. No two x_i or two y_i are equal and no x_i or y_i is zero. Under what further conditions, concerning the x's and the y's, can the expressions ϕ_t and ψ_t then have the same numerical values for all integral values of t, i.e.

$$(H_1 - \overline{H}_1) + \sum_{i=1}^{r} P_i(t)\, x_i^t - \sum_{i=1}^{s} \overline{P}_i(t)\, y_i^t = 0 \quad \text{for} \quad t = 0, 1, 2, \ldots, \tag{6}$$

where $$\phi_t = H_1 + \sum_{i=1}^{r} P_i(t)\, x_i^t \quad \text{with} \quad P_i(t) = \alpha_{i1} + \alpha_{i2} t + \ldots + \alpha_{ik_i} t^{k_i - 1}$$

and $$\psi_t = \overline{H}_1 + \sum_{i=1}^{s} \overline{P}_i(t)\, y_i^t \quad \text{with} \quad \overline{P}_i(t) = \beta_{i1} + \beta_{i2} t + \ldots + \beta_{ij_i} t^{j_i - 1}?$$

Suppose first that none of the x's equals any of the y's. Then it is known that the determinant of any set of equations of the system (6) is not zero. It follows that we must have $H_1 = \overline{H}_1$ and all α's and β's $= 0$, hence all $P_i(t)$ and $\overline{P}_i(t) \equiv 0$. In this case the two populations must already be stationary and therefore identical with the intrinsic stationary populations which are implied by the sets p_i and $\overline{p}_i$, respectively.

On the other hand, if some of the x's are equal to some of the y's, say $x_1 = y_1, \ldots, x_m = y_m$ and all the others are different, then we find by the same argument that $H_1 = \overline{H}_1$ and $P_i = \overline{P}_i$ for the first m values of i, whereas all the other P_i and $\overline{P}_i$ are identically zero. (It is, of course, again possible that *all* the P_i and $\overline{P}_i$ are identically zero and that we have, in fact, again the two intrinsic stationary populations.)

If all x's are equal to the y's, with equal multiplicities, then the two equations are equal

* It follows, for example, from results of P. H. Leslie (1945).

and the two populations must be identical, if they are to be kept constant by equal numbers of entrants.

We have thus reached the following conclusion: If we assume that the two equations given above are not identical, and that the initial populations are not the intrinsic stationary ones, then they must be such that the two equations have some (but not all) roots equal and if we calculate the corresponding P_t and $\overline{P}_t$ (see (4) and (5) of the previous section), then those corresponding to the equal roots must be identical and the others must vanish. This includes the case where the x's and y's are the same, but with different multiplicities, so that the P_t and $\overline{P}_t$ do not all extend to the highest power of t which would be admissible by (7) or (8) respectively.

If the two populations are the intrinsic stationary ones, then the numbers of entrants will be constant (i.e. independent of the year) and the two constants will be equal if and only if

$$\frac{\Sigma l_t}{\Sigma p_t} = \frac{\Sigma \bar{l}_t}{\Sigma \bar{p}_t},$$

where the first expression refers to the first and the second expression to the second population.

Example. In the example used in § 1 we have $\alpha_3 = \alpha_4 = 0$, and we can therefore try to obtain a second population which is kept constant by the same numbers of entrants as the first one. We construct an equation which has again the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$ and $-\frac{1}{3}$, but not $-\frac{1}{8} \pm \sqrt{-\frac{7}{64}}$. Such an equation is, for example,

$$480x^5 + 396x^4 + 218x^3 + 62x^2 + 13x + 1 = 0,$$

which has the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$, $-\frac{1}{3}$, and $-\frac{1}{10} \pm \sqrt{-\frac{9}{100}}$. It implies that the probabilities of survival, i.e. $\overline{p}_t$, are

$$33/40 \quad 109/240 \quad 31/240 \quad 13/480 \quad 1/480,$$

and as the ϕ_t (and the $r_t = \phi_{-t}$) are to be the same as in § 1, the initial population must now be $\bar{l}_t = r_t \overline{p}_t$. Table 2 shows the development of such a population, and it will be seen that the first line is identical with that in Table 1.

Table 2

t	0	1	2	3	4	5	6	7	8 and after
Age 0	859	996	1025	989	1001	1001	999	1000	1000
1	1196	709	822	845	816	826	826	824	825 (= 1000 × 33/40)
2	204	658	390	452	465	449	455	455	454 (= 1000 × 109/240)
3	41	58	187	111	129	132	128	129	129 (= 1000 × 31/240)
4	92	9	12	39	23	27	27	27	27 (= 1000 × 13/480)
5	45	7	1	1	3	2	2	2	2 (= 1000 × 1/480)
	2437	2437	2437	2437	2437	2437	2437	2437	2437

3. STRATIFIED POPULATION: TWO GRADES

The results of the previous sections will now be used for an investigation of the stratified population.* First, we consider a population split into a lower and a higher grade in the following way:

We assume that all members of age 0 are in the lower grade only, but that at all other ages may share in both grades. Apart from mortality, which operates on all members according to their age, we assume that at every age a certain proportion dependent on that age is 'promoted', at the end of the year, from the lower into the higher grade. Our problem is to discover whether this can be done whilst maintaining the totals in both grades constant; naturally the grand total of the population must remain constant.

It is sufficient to deal only with the lower grade, as the numbers at each age in the higher one can be found by subtracting those in the lower grade from the total population at that age. Now the lower grade is depleted by mortality and also by promotions. If the probability of remaining unpromoted until age i is t_i, then the probability of not leaving the grade in this period is $p_i t_i = \overline{p}_i$, say. Since all entrants into the population are at the same time entrants into the lower grade, our problem thus reduces to the following:

Is it possible to find an initial population, stratified into two grades, such that, on the basis of mortality described by p_i, the number of entrants every year necessary to keep the population constant is the same as that calculated on the basis of mortality-cum-promotion, described by $\overline{p}_i$?

We can apply our results in § 2 to this case by considering the lower grade and the total population as the two populations given. It follows that the lower grade can only be kept constant by that number of entrants which is necessary for the total population, if the latter is initially such that some of the $P_i(t)$ which depend on it are either identically zero or at least do not extend to the highest degree indicated by the multiplicities of the corresponding roots in $\Sigma p_i x^{n-i} = 0$. In order to find a suitable initial population for the lower grade it is then necessary to find an equation $\Sigma \overline{p}_i y^{n-i} = 0$ which has the roots, with the necessary multiplicities, which appear explicitly in ϕ_t as calculated from the original equation, but which is not identical with it. The degree of $\Sigma \overline{p}_i y^{n-i} = 0$ may be lower than or equal to that of $\Sigma p_i x^{n-i} = 0$. If it is lower, then all members of the population will be in the higher grade at the highest age or ages.

This condition is not sufficient, however. In view of the interpretation of the equation containing the $\overline{p}_i$'s these coefficients must be positive and, as the lower grade is a part of the whole, we must have $\overline{p}_i \leqslant p_i$ for all i. But it is not necessary that we have also $\overline{p}_{i+1} \leqslant \overline{p}_i$. If the opposite holds, this could still bear a practical interpretation. It would mean that reversions occur from the higher into the lower grade.

If an equation with the necessary and sufficient properties can be found, then we take the $r_i = l_i/p_i$ which we had to start with and construct the initial population of the lower grade by writing the number at age i as $\overline{l}_i = r_i \overline{p}_i = l_i \overline{p}_i / p_i$.

It will be seen that in such a population the age distributions change with the passage of time (tending to a stationary limit) but that nevertheless all entrants have the same combined prospects of survival and promotion. (Thus from the point of view of a member of the community his position is the same as if he entered a stationary population. His chances

* Cf., for the stationary case, with continuous changes, H. L. Seal (1945).

of promotion are unaffected by the changes in the age distribution of those in front of him. But the characteristics of the population as a whole, for instance the efficiency of the staff from the point of view of an employer may, of course, vary considerably.) Such a population will be called *semi-stationary*.

Example. The population shown in Table 2 can be taken as representing a lower grade within the population given in Table 1. The ratios $t_i = \overline{p}_i/p_i$ are then:

t_0	t_1	t_2	t_3	t_4	t_5
1	33/35	218/245	62/75	4/5	4/5

Table 3 is constructed by subtracting Table 2 from Table 1 and thus shows the composition of the higher grade.

Table 3

t	0	1	2	3	4	5	6	7	8 and after
Age 1	73	43	50	52	49	50	50	50	50
2	25	82	48	56	58	56	56	56	56
3	9	12	40	23	27	28	27	27	27
4	23	2	3	10	6	7	7	7	7
5	11	2	—	—	1	—	1	1	1
	141	141	141	141	141	141	141	141	141

4. Stratified population: more than two grades

Let us now split up the higher grade as well. We have then, say, k grades, with grades 2 and above forming the aggregate which was simply called the higher grade in § 2; grade 1 is identical with the lower grade of that section.

We assume further that promotions from any grade into the next higher one take place at the end of every year and that every promotee into any grade has to stay there for at least one year. Thus in any population the lowest possible age of grade g is $g-1$. The actual lowest ages may be different, because the first promotion rates different from 0 may concern higher ages than these. The rates of promotion can be different from grade to grade, but depend within each grade only on the age, as before.

We shall again investigate whether it is possible to keep the total numbers of every grade constant, even if the age distributions of the grades are changing.

We have seen that the age distribution of the total population, after t years, is

$$\phi_t, \quad \phi_{t-1}p_1, \quad \ldots, \quad \phi_{t-n}p_n.$$

The distribution of grade 1 is, at the same time,

$$\phi_t, \quad \phi_{t-1}\overline{p}_1, \quad \ldots, \quad \phi_{t-n}\overline{p}_n,$$

and it is assumed that the set of p_i is not identical with the set of $\overline{p}_i$. Hence grades 2 and above will have the age distribution

$$\phi_t(p_0-\overline{p}_0), \quad \phi_{t-1}(p_1-\overline{p}_1), \quad \phi_{t-n}(p_n-\overline{p}_n).$$

Let us assume that $\phi_{t-\nu}(p_\nu-\overline{p}_\nu)$ is the first item in this series which is not zero. Clearly we have $\nu \geqslant 1$. Then, as far as numbers of members (and not their individual careers) are con-

cerned, this aggregate of grades 2 and above is equivalent to a population which has arisen from successive annual entrants $\phi_{t-\nu}(p_\nu - \overline{p}_\nu)$ who have been subject to rates of survival

$$q_1 = \frac{p_{\nu+1} - \overline{p}_{\nu+1}}{p_\nu - \overline{p}_\nu}, \quad \ldots, \quad q_{n-\nu} = \frac{p_n - \overline{p}_n}{p_\nu - \overline{p}_\nu}$$

It must be understood, however, that 'survival' is here a balance between deaths and promotions into the grade, so that these rates may very well exceed unity.

The number of annual entrants into grade 2 is given by

$$\phi_{t-\nu}(p_\nu - \overline{p}_\nu) = \left[H_1 + \sum_{i=1}^{m} P_i(t-\nu) x_i^{t-\nu} \right] (p_\nu - \overline{p}_\nu),$$

where $x_1, \ldots, x_m$ are the common roots of $\sum_{i=0}^{n} p_i x^{n-i} = 0$ and $\sum_{i=1}^{n} \overline{p}_i y^{n-i} = 0$, with multiplicities k_i and j_i respectively, and where the P_i are polynomials whose order does not exceed either $k_i - 1$ or $j_i - 1$. (They may all be identically zero.)

The x_i are, of course, also roots of $\sum_{i=\nu}^{n} (p_i - \overline{p}_i) x^{n-i} = 0$, with multiplicities given by the smaller of k_i and j_i.

We ask now if it is possible to construct grade 2 alone in such a way that its total remains also constant. The argument which has been used in § 3 shows that this is possible if another equation of grade $n - \nu$ can be found whose coefficients w_i, say, are not larger than the corresponding q_i (and $w_0 = 1$), which has once again the roots $x_1, \ldots, x_m$, with multiplicities g_i at least. If $g_1 + \ldots + g_m = n - \nu$, then this is clearly impossible. If $g_1 + \ldots + g_m$ is smaller than this value, then we can try to find such an equation. The initial population can also be then found, if we multiply the initial population of grades 2 and above by w_i/q_i. The grades 1, 2 and the aggregate of 3 and above can then be constructed and every stratum kept constant, but with changing age distributions.

We can proceed in the same way and find at each step whether further splitting up is possible beyond 3 grades, 4 grades, etc. It is seen that in general, if $\Sigma g_i = n - m$, and if grade g starts in fact at age $g - 1$, then $m + 1$ grades can exist.

The smallest value of Σg_i is 1, and in this extreme case n grades can be constructed, i.e. one less than the number of ages. The nth grade will then contain the ages $n - 1$ and n. Further, since x_1 is a root of $x - x_1 = 0$, the age distribution of this highest grade is

$$H_1 + \alpha_1 x_1^{t-n+1}, \quad -(H_1 + \alpha_1 x_1^{t-n}) x_1 = -H_1 x_1 - \alpha_1 x_1^{t-n+1}$$

(x_1 is, of course, negative).

Example. We use again the same example as before. The characteristic equation for the whole population was

$$x^5 + \tfrac{7}{8}x^4 + \tfrac{49}{96}x^3 + \tfrac{5}{32}x^2 + \tfrac{13}{384}x + \tfrac{1}{384} = 0,$$

and that for grade 1 alone

$$x^5 + \tfrac{33}{40}x^4 + \tfrac{109}{240}x^3 + \tfrac{31}{240}x^2 + \tfrac{13}{480}x + \tfrac{1}{480} = 0.$$

The difference between these two equations gives the equation for grade 2 and above

$$x^4 + \tfrac{9}{8}x^3 + \tfrac{13}{24}x^2 + \tfrac{13}{96}x + \tfrac{1}{96} = 0.$$

This equation has, of course, the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$, $-\frac{1}{3}$ which are common to the two characteristic equations of the fifth degree, and also a further root $-\frac{1}{2}$. Now there is

a biquadratic equation with the three specified common roots and not larger coefficients (and having the coefficient of x^4 equal to unity), viz.

$$x^4+\tfrac{33}{40}x^3+\tfrac{17}{48}x^2+\tfrac{1}{15}x+\tfrac{1}{240}=0.$$

The fourth, irrelevant, root is $-1/5$. This equation leads to the following development:

Grade 2 only

t	0	1	2	3	4	5	6	7	8 and after
Age 1	73	43	50	52	49	50	50	50	50
2	18	60	35	41	43	40	41	41	41
3	6	8	26	14	18	19	18	18	18
4	11	—	1	5	2	3	3	3	3
5	4	1	—	—	—	—	—	—	—
	112	112	112	112	112	112	112	112	112

Grade 3 and above

Age 2	7	22	13	15	15	16	15	15	15
3	3	4	14	9	9	9	9	9	9
4	12	2	2	5	4	4	4	4	4
5	7	1	—	—	1	—	1	1	1
	29	29	29	29	29	29	29	29	29

Analysis into further grades is impossible in this case, because the characteristic equation of the third grade does not have any roots apart from the three common roots of all previous equations.

5. Promotion rates dependent on seniority

We still consider more than two grades, but now we will assume that the promotion rates do not depend on the attained age but on the seniority, i.e. on the time spent in the grade, instead. In the lowest grade seniority is equivalent to age, because all members were supposed to enter at the lowest age only. If we consider again the two grades of § 3, but this time take note of differences in seniority, we find the following pattern:

Age	Lower grade	Higher grade				Total
		Seniority				
		0	1		$x-1$	
0	ϕ_t					ϕ_t
1	$\phi_{t-1}p_1t_1$	$\phi_{t-1}p_1(t_0-t_1)$				$\phi_{t-1}p_1$
2	$\phi_{t-2}p_2t_2$	$\phi_{t-2}p_2(t_1-t_2)$	$\phi_{t-2}p_2(t_0-t_1)$			$\phi_{t-2}p_2$
...	.....	.....	.....	...	.....	.....
x	$\phi_{t-x}p_xt_x$	$\phi_{t-x}p_x(t_{x-1}-t_x)$	$\phi_{t-x}p_x(t_{x-2}-t_{x-1})$	...	$\phi_{t-x}p_x(t_0-t_1)$	$\phi_{t-x}p_x$
...	.....	.....	.....	...	.....	.....

Note. $t_i=\bar{p}_i/p_i$ and hence $t_0=1$.

If we consider now promotion from grade 2 into grade 3, and if we introduce u_s, the probability of not being promoted during s years from grade 2 ($u_0 = 1$), we see that grades 2 and 3 (including higher grades, if any) will have the following constitution:

Grade 2

Age	Seniority 0	1		$x-1$
1	$\phi_{t-1}p_1(t_0-t_1)\,u_0$			
2	$\phi_{t-2}p_2(t_1-t_2)\,u_0$	$\phi_{t-2}p_2(t_0-t_1)\,u_1$	...	.....
...	.....	.....	...	.....
x	$\phi_{t-x}p_x(t_{x-1}-t_x)\,u_0$	$\phi_{t-x}p_x(t_{x-2}-t_{x-1})\,u_1$	...	$\phi_{t-x}p_x(t_0-t_1)\,u_{x-1}$
...	.....	.....	...	.....

Grade 3

Age	Seniority 0	1		$x-2$
2	$\phi_{t-2}p_2(t_0-t_1)\,(u_0-u_1)$			
3	$\phi_{t-3}p_3[(t_1-t_2).(u_0-u_1)$ $+(t_0-t_1)\,(u_1-u_2)]$	$\phi_{t-3}p_3(t_0-t_1)\,(u_0-u_1)$		
...	.....	.....	...	.....
x	$\phi_{t-x}p_x[(t_{x-2}-t_{x-1})\,(u_0-u_1)$ $+(t_{x-3}-t_{x-2})\,(u_1-u_2)+\ldots$ $+(t_0-t_1)\,(u_{x-2}-u_{x-1})]$	$\phi_{t-x}p_x[(t_{x-3}-t_{x-2})\,(u_0-u_1)$ $+(t_0-t_1)\,(u_{x-3}-u_{x-2})]$	...	$\phi_{t-x}p_x(t_0-t_1)\,(u_0-u_1)$
...	.....	.....	...	.....

It follows by means of the same argument as before that grade 2 can be kept constant if we can find the u_i such that the equation

$$x^{n-1}p_1(t_0-t_1)+x^{n-2}p_2[(t_1-t_2)+(t_0-t_1)\,u_1]+\ldots$$
$$+p_n[(t_{n-1}-t_n)+(t_{n-2}-t_{n-1})\,u_1+\ldots+(t_0-t_1)\,u_{n-1}]=0$$

has the same roots which were common to $x^n+x^{n-1}p_1+\ldots+p_n=0$ and

$$x^{n-1}p_1(t_0-t_1)+x^{n-2}p_2(t_0-t_2)+\ldots+p_n(t_0-t_n)=0,$$

which is identical with the difference of the first two equations of degree n, referring respectively to the whole population and to the lowest grade. We must further insist that all u_i must have non-negative values, not larger than 1. The coefficients of the powers of x must also be positive, but it is not necessary that $u_{i+1} \leqslant u_i$, unless we do not admit reversions. If m is the number of common roots, then it follows again as in the last section that $n-m+1$ grades could exist which remain constant under the operation of promotions, but that their age and seniority distributions change.

Example. Dealing once more with the same example as in the previous sections, we have to find a biquadratic equation

$$\tfrac{7}{8}(1-\tfrac{33}{35})\,x^4+\tfrac{49}{96}[(\tfrac{33}{35}-\tfrac{218}{245})+(1-\tfrac{33}{35})\,u_1]\,x^3+\tfrac{5}{32}[(\tfrac{218}{245}-\tfrac{62}{75})+(\tfrac{33}{35}-\tfrac{218}{245})\,u_1+(1-\tfrac{33}{35})\,u_2]\,x^2$$
$$+\tfrac{13}{384}[(\tfrac{62}{75}-\tfrac{4}{5})+(\tfrac{218}{245}-\tfrac{62}{75})\,u_1+(\tfrac{33}{35}-\tfrac{218}{245})\,u_2+(1-\tfrac{33}{35})\,u_3]\,x$$
$$+\tfrac{1}{384}[(\tfrac{4}{5}-\tfrac{4}{5})+(\tfrac{62}{75}-\tfrac{4}{5})\,u_1+(\tfrac{218}{245}-\tfrac{62}{75})\,u_2+(\tfrac{33}{35}-\tfrac{218}{245})\,u_3+(1-\tfrac{33}{35})\,u_4]=0,$$

or, if we use four significant figures in every fraction,

$$x^4 + (0{\cdot}5417 + 0{\cdot}5833u_1)x^3 + (0{\cdot}1973 + 0{\cdot}1658u_1 + 0{\cdot}1786u_2)x^2$$
$$+ (0{\cdot}01805 + 0{\cdot}04274u_1 + 0{\cdot}03593u_2 + 0{\cdot}03869u_3)x$$
$$+ (0 + 0{\cdot}001389u_1 + 0{\cdot}003288u_2 + 0{\cdot}002764u_3 + 0{\cdot}002976u_4) = 0.$$

This biquadratic equation must have the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$ and $-\frac{1}{3}$. If the fourth root is called $(-z)$, then the equation must be identical with

$$(x^3 + \tfrac{5}{3}x^2 + \tfrac{11}{48}x + \tfrac{1}{48})(x + z) = 0.$$

Simple arithmetic shows then that

$$u_1 = 0{\cdot}1429 + 1{\cdot}7142z, \quad u_2 = 0{\cdot}0459 + 1{\cdot}9082z,$$
$$u_3 = -0{\cdot}1283 + 2{\cdot}2566z \quad \text{and} \quad u_4 = 0{\cdot}0019 + 1{\cdot}9962z.$$

Now z must be at least 0·05685 to make u_3 positive and it must not exceed 0·5, because otherwise the u_i would exceed unity. But then u_4 will always be larger than u_3, unless we put $z = \frac{1}{2}$ which would mean $u_i = 1$ for all i and then there would be no members at all in grades 3 and above. It follows that we must admit reversions from grade 3 into grade 2. We can then, for instance, take $z = 0{\cdot}2$ and have

$$u_1 = 0{\cdot}4857, \quad u_2 = 0{\cdot}4275, \quad u_3 = 0{\cdot}3230 \quad \text{and finally} \quad u_4 = 0{\cdot}4011.$$

The biquadratic equation becomes

$$x^4 + \tfrac{33}{40}x^3 + \tfrac{17}{48}x^2 + \tfrac{1}{15}x + \tfrac{1}{240} = 0.$$

This is the same as the one used in § 4, and we can again write down the changing pattern of the population, but this time taking also seniority into account:

Grade 2

Age	$t=0$						1					2					3				
1	73	—	—	—	—	73	43	—	—	—	43	50	—	—	—	50	52	—	—	—	52
2	12	6	—	—	—	18	40	20	—	—	60	23	12	—	—	35	27	14	—	—	41
3	3	2	1	—	—	6	4	2	2	—	8	15	6	5	—	26	8	3	3	—	14
4	3	3	3	2	—	11	—	—	—	—	0	—	1	—	—	1	1	2	1	1	5
5	—	—	2	1	1	4	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—
	91	11	6	3	1	112	87	22	3	—	112	88	19	5	—	112	88	19	4	1	112
	4						5					6					7 and later				
1	49	—	—	—	—	49	50	—	—	—	50	50	—	—	—	50	50	—	—	—	50
2	28	15	—	—	—	43	27	13	—	—	40	27	14	—	—	41	27	14	—	—	41
3	10	4	4	—	—	18	10	5	4	—	19	10	4	4	—	18	10	4	4	—	18
4	1	1	—	—	—	2	1	1	1	—	3	1	1	1	—	3	1	1	1	—	3
5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	88	20	4	—	—	112	88	19	5	—	112	88	19	5	—	112	88	19	5	—	112

Grade 3

Age	0						1					2					3				
2	7	—	—	—	—	7	22	—	—	—	22	13	—	—	—	13	15	—	—	—	15
3	1	2	—	—	—	3	2	2	—	—	4	6	8	—	—	14	4	5	—	—	9
4	4	3	5	—	—	12	1	—	1	—	2	—	1	1	—	2	1	2	2	—	5
5	1	2	2	2	—	7	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—
	13	7	7	2	—	29	25	2	1	1	29	19	9	1	—	29	20	7	2	—	29
	4						5					6					7 and later				
2	15	—	—	—	—	15	16	—	—	—	16	15	—	—	—	15	15	—	—	—	15
3	4	5	—	—	—	9	4	5	—	—	9	4	5	—	—	9	4	5	—	—	9
4	1	1	2	—	—	4	1	1	2	—	4	1	1	2	—	4	1	1	2	—	4
5	—	1	—	—	—	1	—	—	—	—	—	—	1	—	—	1	—	1	—	—	1
	20	7	2	—	—	29	21	6	2	—	29	20	7	2	—	29	20	7	2	—	29

We find, as before, that further splitting up of grades is impossible, if the total in each grade is to remain constant throughout the years.

SUMMARY

This investigation deals with a stratified population, which is subject to (i) mortality, dependent on age, and to (ii) promotion rates, indicating the ratios of members of a grade which are transferred to the next higher grade at the end of the year.

Section 1 concerns a population which is not yet stratified and formulae are deduced to calculate the number of entrants at time t, necessary to replace yearly deaths and thus to keep the total of the population constant. This number depends clearly on the mortality rates and on the age distribution existing at time $t = 0$. In general the population tends towards a limiting age distribution, the 'intrinsic stationary population'.

Section 2 considers two populations and conditions are derived for the case that they need, every year, equal numbers of entrants to keep them constant.

Section 3 introduces the stratified population. Both mortality and promotion rates depend on the age, and they are independent of the time t. Under certain conditions one of the two populations considered in § 2 can be taken as the whole and the other as the lowest grade in it. It is shown how and when entries into the grade can, at the same time, replace both losses due to mortality in the whole population, and to mortality and promotion depleting the lowest grade. This can also be described by saying that the totals of both grades can be kept constant at the same time, although the age distributions change from year to year.

Section 4 generalizes the results of the previous section for a population consisting of k grades. If the population is spread over n ages, then it is shown that up to $n-1$ grades

may be possible in the most favourable case, such that they are all kept constant, whilst the age distributions all oscillate. Such a population is called semi-stationary.

Section 5 introduces the case which has been of actual importance in practical establishment work: the promotion rates are made dependent on the time spent in the grade instead of on the age.

A numerical example is attached to § 1 and is carried through all stages to illustrate the results which emerge gradually in the subsequent sections.

REFERENCES

LESLIE, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183.
SEAL, H. L. (1945). The mathematics of a population composed of k stationary strata. *Biometrika*, **33**, 226.

A SIMPLE APPROACH TO CONFOUNDING AND FRACTIONAL REPLICATION IN FACTORIAL EXPERIMENTS

By O. KEMPTHORNE, *Rothamsted Experimental Station*

Introduction

The design and analysis of factorial experiments was described in 1937 by Yates in considerable detail. In his treatment Yates described first the 2^n system and then went on to deal with 3^n experiments and experiments of the 2^m3^n type. The 2^n system is capable of very easy explanation, but with experiments of higher order both the design and analysis become of increasing complexity. It is the purpose of this paper to present a general method by which factorial designs of the type p^n may be examined, in respect of both confounding and fractional replication. The method will be described by explanation of the rules for the 2^n and 3^n systems and corresponds quite closely to that given by Fisher (1942). The present approach presents confounding and fractional replication as different aspects of the same process. Experimental designs suggested by Plackett & Burman (1946) are also discussed.

The 2^n system

In this system all combinations of n factors each at two levels are tested. The totality of treatment combinations may be represented by the points of an n-dimensional lattice, each side being of unit length. Let the factors be $x_1, x_2, \ldots, x_n$ and take n mutually orthogonal axes $y_1 \ldots y_n$. The point $(000 \ldots 0)$ will then represent the control treatment, $(1000 \ldots 0)$ the treatment consisting of x_1 at the upper level and all the other factors at the lower level, and so on. The treatment effect of x_1 is the difference of the means of the yields of plots receiving x_1 and those not receiving x_1. It is therefore the difference between the mean of the plots represented by points lying on the plane $y_1 = 1$ and the mean of those represented by the points on the plane $y_1 = 0$. The interaction of x_1 and x_2 is the difference between the means of those plots represented by $y_1 = 1$, $y_2 = 1$ or $y_1 = 0$, $y_2 = 0$ and those represented by $y_1 = 0$, $y_2 = 1$ and $y_1 = 1$, $y_2 = 0$, i.e. the difference of the means of those plots for which

$$y_1 + y_2 = 2 \text{ or } = 0 \pmod 2,$$

and those for which

$$y_1 + y_2 = 1 \pmod 2.$$

Similarly, the triple interaction of x_1, x_2 and x_3 is the difference between the means of those plots for which

$$y_1 + y_2 + y_3 = 0 \pmod 2,$$

and those for which

$$y_1 + y_2 + y_3 = 1 \pmod 2.$$

This process can be continued to the consideration of the interaction of $x_1, x_2, \ldots, x_n$ which is the difference between the mean of those plots for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 0 \pmod 2,$$

and the mean of those for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 1 \pmod 2.$$

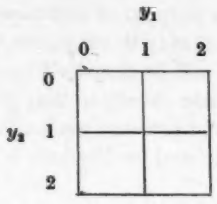
In the n-dimensional space parallel hyper-planes may be drawn containing the points of the lattice, such that the total yield forming the positive part of an interaction is obtained from

a set of parallel hyper-planes equidistant from each other. Likewise the negative part is obtained from another set of parallel hyper-planes, each plane of which lies midway between two planes of the first set.

The 3^n system

With n factors at each of three levels the treatment combinations are given by an n-dimensional lattice, each side being of length two units and containing three points. The treatment contrasts may be described as in the 2^n system with some slight modifications.

Any contrast in the 3^n system involves the comparison of three totals of the yields of 3^{n-1} plots, and may be represented by the comparison of the differences between the yields of the plots lying on three sets of parallel hyper-planes. For example, if $n = 2$ the lattice is as follows:



The main effect of x_1 is the difference between the totals of yields of plots for which $y_1 = 0$, $y_1 = 1$ and $y_1 = 2$. The I component* of the interaction of x_1 and x_2 is the difference between the totals of the yields of plots for which

$$y_1 - y_2 = 0, \quad y_1 - y_2 = 1, \quad \text{and} \quad y_1 - y_2 = 2.$$

The J component is given by the contrast between the yields of plots for which

$$y_1 + y_2 = 0, \quad y_1 + y_2 = 1, \quad \text{and} \quad y_1 + y_2 = 2.$$

Anticipating the extension to cases when n is greater than 2, the equations for the I component may be written as follows:

$$X_1X_2(I_0): \quad y_1 + 2y_2 = 0 \pmod 3,$$
$$X_1X_2(I_1): \quad y_1 + 2y_2 = 1 \pmod 3,$$
$$X_1X_2(I_2): \quad y_1 + 2y_2 = 2 \pmod 3.$$

If x_1 and x_2 (and therefore y_1 and y_2) are interchanged, then $X_2X_1(I_0)$ is given by the equations $y_2 + 2y_1 = 0$, $X_2X_1(I_1)$ by $y_2 + 2y_1 = 1$, $X_2X_1(I_2)$ by $y_2 + 2y_1 = 2$, all mod 3. But the equation $y_2 + 2y_1 = 0 \pmod 3$ is identical with the equation $y_1 + 2y_2 = 0 \pmod 3$, since $3y_1 + 3y_2 = 0 \pmod 3$, whatever the values of y_1 and y_2; $X_2X_1(I_0)$ is therefore equal to $X_1X_2(I_0)$. Subtracting the equation $y_2 + 2y_1 = 1 \pmod 3$ from the equation

$$3y_1 + 3y_2 = 0 = 3 \pmod 3,$$

we get $y_1 + 2y_2 = 2 \pmod 3$; $X_2X_1(I_1)$ is therefore identical with $X_1X_2(I_2)$. It is obvious from the equations given above for the J component that $X_1X_2(J_i) = X_2X_1(J_i)$ for $i = 0$, 1 and 2.

* Yates's terminology for the components of interactions is used where convenient, but it is more convenient to refer to I_1, I_2 and I_3 as I_0, I_1 and I_2 respectively.

Considering the case $n = 3$, it is easily seen that the second order interaction may be split into four parts each consisting of the contrasts between three totals. These may be represented by the following equations:

$$\begin{aligned}
&\text{(I)} && y_1 + y_2 + y_3 = 0 \pmod 3,\\
&&& y_1 + y_2 + y_3 = 1\\
&&& y_1 + y_2 + y_3 = 2\\
&\text{(II)} && y_1 + 2y_2 + y_3 = 0 \pmod 3,\\
&&& y_1 + 2y_2 + y_3 = 1\\
&&& y_1 + 2y_2 + y_3 = 2\\
&\text{(III)} && y_1 + y_2 + 2y_3 = 0 \pmod 3,\\
&&& y_1 + y_2 + 2y_3 = 1\\
&&& y_1 + y_2 + 2y_3 = 2\\
&\text{(IV)} && y_1 + 2y_2 + 2y_3 = 0 \pmod 3.\\
&&& y_1 + 2y_2 + 2y_3 = 1\\
&&& y_1 + 2y_2 + 2y_3 = 2
\end{aligned}$$

In order these have been named by Yates

$$Z, \quad X, \quad Y, \quad W.$$

It is interesting in passing to note the relations between Z, X, Y and W for permutations of the order of the factors. It is obvious from (I) that Z is invariant for any change in order of the three factors X_1, X_2 and X_3. Interchanging y_2 and y_3, equations (II) become equations (III), so that $ABC(X) = ACB(Y)$. The following interchanges may be easily verified (using the equation $3y_1 + 3y_2 + 3y_3 = 0 \pmod 3$ where necessary):

$$ABC(X) = BCA(Y) = CAB(Y) = ACB(Y) = CBA(X) = BAC(W).$$

From the equations, it is clear that Z, X and W may be computed in the way given by Yates, since

$$Z = J\{x_1, J(x_2, x_3)\}, \qquad Y = J\{x_1, I(x_2, x_3)\},$$
$$X = I\{x_1, I(x_2, x_3)\}, \qquad W = I\{x_1, J(x_2, x_3)\},$$

$I(x_2, x_3)$ and $J(x_2, x_3)$ being evaluated for each level of x_1. The extension to the case $n = 4$ is again obvious; the main effects, two-factor and three-factor interactions, follow as in the above, and the four-factor interaction may be split into eight comparisons of three totals:

$$\begin{aligned}
&\text{I} && y_1 + y_2 + y_3 + y_4 = 0, 1, 2 \pmod 3,\\
&\text{II} && y_1 + y_2 + y_3 + 2y_4 = 0, 1, 2 \pmod 3,\\
&\text{III} && y_1 + y_2 + 2y_3 + y_4 = 0, 1, 2 \pmod 3,\\
&\text{IV} && y_1 + y_2 + 2y_3 + 2y_4 = 0, 1, 2 \pmod 3,\\
&\text{V} && y_1 + 2y_2 + y_3 + y_4 = 0, 1, 2 \pmod 3,\\
&\text{VI} && y_1 + 2y_2 + y_3 + 2y_4 = 0, 1, 2 \pmod 3,\\
&\text{VII} && y_1 + 2y_2 + 2y_3 + y_4 = 0, 1, 2 \pmod 3,\\
&\text{VIII} && y_1 + 2y_2 + 2y_3 + 2y_4 = 0, 1, 2 \pmod 3.
\end{aligned}$$

As in the case of two factors, the effect of permutations of the order on the components of X, Y, Z and W may be easily obtained. The four-factor interactions may be computed by putting the equations given above into the following form:

$$\begin{aligned} \text{I} &= J\{x_1, Z\}, & \text{V} &= I\{x_1, W\}, \\ \text{II} &= J\{x_1, Y\}, & \text{VI} &= I\{x_1, X\}, \\ \text{III} &= J\{x_1, X\}, & \text{VII} &= I\{x_1, Y\}, \\ \text{IV} &= J\{x_1, W\}, & \text{VIII} &= I\{x_1, Z\}, \end{aligned}$$

where the three components of W, X, Y, Z of x_2, x_3, x_4 (in that order) are evaluated for each level of x_1.

The p^n system

The total of p^n-1 degrees of freedom, where p is a prime, in the analysis of variance of a p^n experiment may be split into $(p^n-1)/(p-1)$ sets of $(p-1)$ degrees of freedom, the contrasts being given by the following hyper-planes:

$$\text{Main effects} \quad \left.\begin{aligned} &y_1 = 0, 1, 2, \ldots, p-1, \\ &y_2 = 0, 1, 2, \ldots, p-1. \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &y_p = 0, 1, 2, \ldots, p-1. \end{aligned}\right\} \pmod p,$$

$$\text{Interactions of pairs of factors, e.g. of } x_1 \text{ and } x_2 \quad \left\{\begin{aligned} &y_1 + y_2 = 0, 1, 2, \ldots, p-1, \\ &y_1 + 2y_2 = 0, 1, 2, \ldots, p-1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &y_1 + (p-1)y_2 = 0, 1, 2, \ldots, p-1. \end{aligned}\right. \pmod p,$$

and so on to the interaction between all the factors which is given by the hyper-planes

$$a_1 y_1 + a_2 y_2 + a_3 y_3 + \ldots + a_n y_n = 0, 1, 2, \ldots, p-1 \pmod p,$$

where a_1 equals 1 and $a_2, a_3, \ldots, a_n$ each may take all values from 1 to $p-1$.

Simplification of notation

The p^n-1 degrees of freedom in the p^n system may be split into $(p^n-1)/(p-1)$ sets of $(p-1)$ degrees of freedom, given by the above hyper-planes, but it is only necessary to specify one hyper-plane of each set of the parallel hyper-planes.

All the comparisons may be denoted by $y_1^{\alpha_1} y_2^{\alpha_2}, \ldots, y_n^{\alpha_n}$, the symbol meaning that the comparisons are given by the hyper-planes

$$\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_n y_n = 0, 1, 2, \ldots, p-1 \pmod p.$$

In order to obtain an enumeration which covers all the possibilities once and once only, it is necessary to use the rule that the factors are always written down in ascending order—i.e. $y_i^{\alpha_i} y_j^{\alpha_j} y_k^{\alpha_k}$, etc., such that $i < j < k \ldots$ and that $\alpha_i = 1$.

THE 3^n SYSTEM IN THE REVISED NOTATION

As an example, the 3^3 system will be examined in detail. The effects are represented by y_1, y_2, y_3; interactions between pairs, $y_1y_2, y_1y_2^2, y_1y_3, y_1y_3^2, y_2y_3, y_2y_3^2$; interactions between all three factors, $y_1y_2y_3, y_1y_2y_3^2, y_1y_2^2y_3, y_1y_2^2y_3^2$. Any other combination of powers of the y's can be reduced to the above set.

It is interesting to examine the interactions of the effects and interactions. In the case of the 2^n system, Yates refers to the generalized interaction of two interactions $ABCD$ and CDE say, which is ABE. The interaction of effects or interactions A and B consists of AB and AB^2 in the 3^n system.

(*a*) The interactions of main effects are obviously interactions between pairs of factors.

(*b*) The interactions of main effects and two-factor interactions with one letter in common are two-factor interactions and main effects: e.g. the interactions of y_1 and y_1y_2 are

$$y_1^2y_2 = y_1y_2^2, \quad \text{and} \quad y_1^2y_2^2 = y_2,$$

and the interactions of y_1 and $y_1y_2^2$ are $y_1^2y_2^2 = y_1y_2$, and $y_1^3y_2^4 = y_2$.

(*c*) The interaction between main effects and three-factor interaction are two-factor and three-factor interactions:

Between	Interactions
y_1 and $y_1y_2y_3$	$y_1^2y_2y_3 = y_1y_2^2y_3^2$, $y_1^2y_2^2y_3^2 = y_2y_3$
y_1 and $y_1y_2y_3^2$	$y_1^2y_2y_3^2 = y_1y_2^2y_3$, $y_1^2y_2^2y_3^4 = y_2y_3^2$
y_1 and $y_1y_2^2y_3$	$y_1^2y_2^2y_3 = y_1y_2y_3^2$, $y_1^2y_2^4y_3^2 = y_2y_3^2$
y_1 and $y_1y_2^2y_3^2$	$y_1^2y_2^2y_3^2 = y_1y_2y_3$, $y_1^2y_2^4y_3^4 = y_2y_3$

(*d*) The interaction between two-factor interactions are exemplified in the following table:

Between	Interactions
y_1y_2 and $y_1y_2^2$	$y_1^2y_2^2 = y_1$, $y_1^2y_2^4 = y_2$
y_1y_2 and y_2y_3	$y_1y_2^2y_3$, $y_1y_2^2y_3^2 = y_1y_3^2$
y_1y_2 and $y_2y_3^2$	$y_1y_2^2y_3^2$, $y_1y_2^2y_3^4 = y_1y_3$

(*e*) The interaction between two-factor and three-factor interactions are exemplified in the following table:

Between / and	y_1y_2		$y_1y_2^2$	
$y_1y_2y_3$	$y_1y_2y_3^2$	y_3	$y_1y_3^2$	$y_2y_3^2$
$y_1y_2y_3^2$	$y_1y_2y_3$	y_3	y_1y_3	y_2y_3
$y_1y_2^2y_3$	$y_1y_3^2$	y_2y_3	$y_1y_2^2y_3^2$	y_3
$y_1y_2^2y_3^2$	y_1y_3	$y_2y_3^2$	$y_1y_2^2y_3$	y_3

The interactions between two-factor and three-factor interactions are therefore two-factor interactions in some cases and main effects and three-factor interactions in the other cases.

(*f*) The interactions between three-factor interactions are set out in the following diagram:

Between / and	$y_1y_2y_3$	$y_1y_2y_3^2$	$y_1y_2^2y_3$	$y_1y_2^2y_3^2$
$y_1y_2y_3$	— —	y_1y_2, y_3	y_1y_3, y_2	y_1, y_2y_3
$y_1y_2y_3^2$		— —	y_1, $y_2y_3^2$	$y_1y_3^2$, y_2
$y_1y_2^2y_3$			— —	$y_1y_2^2$, y_3
$y_1y_2^2y_3^2$				— —

Confounding

Confounding or the allocation of treatment combinations to blocks implies the allocation of all the points of the lattice into p^c sets, of p^{n-c} points, such that the comparisons between these sets involve particular sets of $p-1$ degrees of freedom. The aim of confounding is to reduce the effect of soil heterogeneity by reducing block size, but ensuring that the block comparisons have little possible practical importance.

If comparisons $A = y_1^{\alpha_1}y_2^{\alpha_2}\dots y_n^{\alpha_n}$ and $B = y_1^{\beta_1}y_2^{\beta_2}\dots y_n^{\beta_n}$ are confounded, then so is their generalized interaction, i.e. all the products of these two, i.e. AB, AB^2, ..., AB^{p-1}. For, if the treatment combinations for which $\alpha_1y_1+\alpha_2y_2+\dots+\alpha_ny_n$ is equal to $0, 1, 2, \dots, p-1$ are put into separate blocks and also those treatments for which $\beta_1y_1+\beta_2y_2+\dots+\beta_ny_n$ is equal to $0, 1, 2, \dots, p-1$, then $(\alpha_1+\lambda\beta_1)y_1+(\alpha_2+\lambda\beta_2)y_2+\dots+(\alpha_n+\lambda\beta_n)y_n$ is equal (mod p) to $0, 1, 2, \dots, p-1$ for all λ from 0 to $p-1$.

The present approach to confounding of the 2^n system is identical with that given by Yates and we proceed to consider the rather more complex case of the 3^n system.

(*a*) 3^3 *system*

(1) *In blocks of* 3^2. Any three-factor interaction may be confounded.

(2) *In blocks of* 3. We cannot confine the confounded degrees of freedom to three-factor interactions because the generalized interaction of any two reduces to a two-factor interaction and a main effect. If two three-factor interactions, $y_1y_2y_3$ and $y_1y_2^2y_3$ are confounded, the 8 degrees of freedom for blocks may be described as follows:

	D.F.
y_2	2
y_1y_3	2
$y_1y_2y_3$	2
$y_1y_2^2y_3$	2
	8

We can, however, choose three two-factor interactions and one three-factor interaction pair for our block comparisons.

(*b*) 3^4 *system in blocks of* 3^2

It is immediately obvious that we can confound two two-factor interactions and two higher-order interaction pairs to give blocks of nine. The important point, however, is to find a design confounding only three-factor interaction pairs.

We therefore evaluate the interactions of all pairs of three-factor interactions, which have two letters in common. These may be derived from the interaction of $y_1y_2y_3$ with the four three-factor interactions of y_1, y_2 and y_4, which are as follows:

Interaction of $y_1y_2y_3$	and	$y_1y_2y_4$	$y_1y_2y_3^2y_4^2$	and	$y_3y_4^2$,
Interaction of $y_1y_2y_3$	and	$y_1y_2y_4^2$	$y_1y_2y_3^2y_4$	and	y_3y_4,
Interaction of $y_1y_2y_3$	and	$y_1y_3^2y_4$	$y_1y_3^2y_4^2$	and	$y_2y_3^2y_4$,
Interaction of $y_1y_2y_3$	and	$y_1y_3^2y_4^2$	$y_1y_3^2y_4$	and	$y_2y_3^2y_4^2$.

Obviously there are many designs for the 3^4 design in nine blocks of nine plots confounding three-factor interactions. Those which confound four-factor interactions must also confound two-factor interactions. The names of the confounded interactions and their squares (each of which corresponds to the same grouping as the element itself) form a group with the identity and the equation $y_i^3 = 1$, for all i, and further work is presumably most promising on these lines.

(*c*) 3^5 *in blocks of* 9

There is no design confounding only three-factor or higher-order interactions. If one two-factor interaction can be sacrificed, a possible scheme of confounding is given by the following table of generalized interactions:

Between / and	$y_1y_2y_3$	$y_1y_3^2y_4^2$	$y_1y_3^2y_4$	$y_2y_3^2y_4^2$
$y_3y_4y_5$	$y_1y_2y_3^2y_4y_5$ $y_1y_2y_4^2y_5^2$	$y_1y_3^2y_2y_5$ $y_1y_3^2y_4^2y_5^2$	$y_1y_3^2y_5$ $y_1y_2y_5^2$	y_2y_5 $y_2y_3y_4y_5^2$

This two-factor interaction is estimated by the comparison of three sets of nine blocks, and the accuracy of the estimate will be low.

(*d*) 3^6 *in blocks of* 27

We may, for example, confound the following:

	$y_3y_4y_6$	
$y_1y_2y_3$	$y_1y_3^2y_2y_4y_6$	$y_1y_2y_3^2y_4^2$
$y_1y_4y_5$	$y_1y_2y_4^2y_5y_6$	$y_1y_3^2y_5y_6^2$
$y_1y_3^2y_4^2y_5^2$	$y_1y_3^2y_5^2y_6$	$y_1y_2y_3^2y_4y_5^2y_6^2$
$y_2y_3y_4^2y_5^2$	$y_2y_3^2y_4y_5^2$	$y_2y_4y_5^2y_6^2$

Three three-factor interactions, six four-factor interactions, three five-factor interactions and one six-factor interaction are confounded. If y_6 is omitted from all the above expressions

we obtain a 3^5 experiment in blocks of nine confounding one two-factor interaction, seven three-factor interactions, three four-factor and two five-factor interactions—that is, the design given above for the 3^5 system.

Extension to more complicated cases

Extensions of the above to more complicated cases should most easily be achieved by the use of group theory. The confounding of a p^n design in p^c blocks corresponds to a group of $\frac{1}{2}(p^c+1)$ elements such that all except the unit element involve at least a certain number of letters. For most agricultural experiments each element should contain at least three letters, so that no main effects or two-factor interactions are confounded. The group is an Abelian group and if A and B are elements of the group so are $AB, AB^2, \ldots, AB^{p-1}$. The order of each element is p, and if A is an element so are the first $(p-1)$ powers of A. This aspect is being followed, and it is hoped will yield results.

Fractional replication in the 2^n system

Some principles of fractional replication have been worked out over the past few years at Rothamsted (Finney, 1945). In the case of a 2^n system, with factors $a_1 \ldots a_n$ say, a half-replicate might consist of those treatment combinations which form the positive part of the interaction $A_1 A_2 \ldots A_n$. Each function of the plot yields consisting of the sum of one-half of them minus the sum of the other half then corresponds to two degrees of freedom. Alternatively, each degree of freedom has one alias, and the aim in fractional replication is to design the experiment so that the aliases of effects which the experimenter wishes to measure are high-order interactions which could not possibly have practical significance.

For convenience of presentation, we develop first the theory for the case of the 2^n system. Suppose that of all the points on the lattice for the 2^n system, only those points for which

$$y_1+y_2+y_3+\ldots+y_n = 0$$

are included in the experiment. Then the points on the hyper-plane $y_1 = 0$, also lie on the plane $y_2+y_3+\ldots+y_n = 0$, and likewise those for which $y_1 = 1$ lie on the plane

$$y_2+y_3+\ldots+y_n = 1.$$

The contrast which we have denoted by y_1 is therefore identical with that denoted by $y_2 y_3 y_4 \ldots y_n$. Again, if we suppose that only those treatment combinations are tested which lie on the hyper-planes

$$\alpha_1 y_1+\alpha_2 y_2+\ldots+\alpha_n y_n = 0, \quad \beta_1 y_1+\beta_2 y_2+\ldots+\beta_n y_n = 0,$$

then the points will also lie on the intersection of these planes which is given by the equation

$$(\alpha_1+\beta_1) y_1+(\alpha_2+\beta_2) y_2+\ldots+(\alpha_n+\beta_n) = 0 \pmod 2.$$

The points which lie on the hyper-planes

$$\gamma_1 y_1+\gamma_2 y_2+\ldots+\gamma_n y_n = 0, 1 \pmod 2$$

will also lie on the planes

$$\begin{aligned}
(\alpha_1+\gamma_1) y_1+(\alpha_2+\gamma_2) y_2 \quad &+\ldots+ \quad (\alpha_n+\gamma_n) y_n = 0, 1 \pmod 2,\\
(\beta_1+\gamma_1) y_1+(\beta_2+\gamma_2) y_2 \quad &+\ldots+ \quad (\beta_n+\gamma_n) y_n = 0, 1 \pmod 2,\\
(\alpha_1+\beta_1+\gamma_1) y_1+(\alpha_2+\beta_2+\gamma_2) y_2+\ldots+&(\alpha_n+\beta_n+\gamma_n) y_n = 0, 1 \pmod 2.
\end{aligned}$$

Changing to the simpler notation, these results may be obtained by equating to unity the symbols corresponding to the effects which the experiment cannot measure (as only treat-

ment combinations of the same sign in the function giving the effect are included) and multiplying the symbol corresponding to a particular effect by these symbols. Thus we put

$$I = y_1^{\alpha_1} y_2^{\alpha_2} \dots y_n^{\alpha_n} = y_1^{\beta_1} y_2^{\beta_2} \dots y_n^{\beta_n} = y_1^{\alpha_1+\beta_1} y_2^{\alpha_2+\beta_2} \dots y_n^{\alpha_n+\beta_n},$$

then the contrast $y_1^{\gamma_1} y_2^{\gamma_2} \dots y_n^{\gamma_n}$

is the same as those given by

$$y_1^{\alpha_1+\gamma_1} y_2^{\alpha_2+\gamma_2} \dots y_n^{\alpha_n+\gamma_n}, \quad y_1^{\beta_1+\gamma_1} y_2^{\beta_2+\gamma_2} \dots y_n^{\beta_n+\gamma_n} \quad \text{and} \quad y_1^{\alpha_1+\beta_1+\gamma_1} y_2^{\alpha_2+\beta_2+\gamma_2} \dots y_n^{\alpha_n+\beta_n+\gamma_n},$$

where each power is reduced modulus 2.

2^n SYSTEM WITHOUT SUBDIVISION INTO BLOCKS

We now consider some of the possibilities of partial replication for the 2^n system. The basis of designs with fractional replication is the choice of an identity relationship; most of the possible relationships are of no value, and we consider only those which yield the least possible confusion between main effects and first-order interactions.

Half-replication

$n = 3$. If we take $I = y_1 y_2 y_3$, then $y_1 = y_1(y_1 y_2 y_3) = y_1^2 y_2 y_3 = y_2 y_3$. Such a design which confuses main effects and two-factor interactions would not be of any practical use.

$n = 4$. If we take $I = y_1 y_2 y_3 y_4$, then the aliases are exemplified by

$$y_1 = y_2 y_3 y_4 \quad \text{and} \quad y_1 y_2 = y_3 y_4.$$

Such a design would not be used unless the experimenter were confident that two-factor interactions were negligible.

$n = 5$. If we take $I = y_1 y_2 y_3 y_4 y_5$, then the aliases are exemplified by

$$y_1 = y_2 y_3 y_4 y_5 \quad \text{and} \quad y_1 y_2 = y_3 y_4 y_5.$$

A half-replicate with five or more factors is feasible when there is no necessity to remove heterogeneity by the use of blocks, since main effects will have aliases which are interactions of four factors at least, and two-factor interactions will have aliases which are interactions of at least three factors.

Quarter-replication

Each degree of freedom will now have three aliases. For each value of n we give the identity relationship and typical alias relationships.

$n = 4$. $I = y_1 y_2 = y_3 y_4 = y_1 y_2 y_3 y_4$;

then $y_1 = y_2 = y_1 y_3 y_4 = y_2 y_3 y_4$ and $y_1 y_3 = y_2 y_3 = y_1 y_4 = y_2 y_4$.

$n = 5$. $I = y_1 y_2 = y_3 y_4 y_5 = y_1 y_2 y_3 y_4 y_5$ (a),

or $I = y_1 y_2 y_3 = y_3 y_4 y_5 = y_1 y_2 y_4 y_5$ (b).

(a) Gives $y_1 = y_2 = y_1 y_3 y_4 y_5 = y_2 y_3 y_4 y_5$ and $y_1 y_3 = y_2 y_3 = y_1 y_4 y_5 = y_2 y_4 y_5$.

(b) Gives $y_1 = y_2 y_3 = y_1 y_3 y_4 y_5 = y_2 y_4 y_5$.

$n = 6$. $I = y_1 y_2 y_3 y_4 = y_3 y_4 y_5 y_6 = y_1 y_2 y_5 y_6$;

then $y_1 = y_2 y_3 y_4 = y_1 y_3 y_4 y_5 y_6 = y_2 y_5 y_6$

and $y_1 y_2 = y_3 y_4 = y_1 y_2 y_3 y_4 y_5 y_6 = y_5 y_6$.

$n = 7$. $I = y_1 y_2 y_3 y_4 = y_4 y_5 y_6 y_7 = y_1 y_2 y_3 y_5 y_6 y_7$;

then $y_1 = y_2 y_3 y_4$ and $y_1 y_2 = y_3 y_4$.

$n = 8$. $I = y_1 y_2 y_3 y_4 y_5 = y_4 y_5 y_6 y_7 y_8 = y_1 y_2 y_3 y_6 y_7 y_8$;

then $y_1 = y_2 y_3 y_4 y_5$ and $y_1 y_2 = y_3 y_4 y_5$.

Designs in quarter replicate are therefore possible when n is greater than or equal to 8.

High-order fractional replication

In general, the existence of fractional designs of the 2^n system with fraction 2^p, which will be useful where information on all main effects and two-factor interactions is required, depends on the existence of a group of 2^p elements, one element being unity and the other elements all containing at least five letters. No simple method has been found of enumerating such groups, but it is perhaps worth recording the following designs which appear to represent the greatest degree of fractional replication possible.

(*a*) *Eighth replication*

If we are testing ten or more factors at each of two levels, one-eighth of a replication will enable main effects and two-factor interactions to be estimated. An appropriate identity relationship is the following:

$$I = y_1y_2y_3y_4y_5 = y_1y_2y_6y_7y_8 = y_3y_4y_5y_6y_7y_8$$
$$= y_1y_3y_7y_9y_{10} = y_2y_4y_5y_7y_9y_{10} = y_2y_3y_6y_8y_9y_{10} = y_1y_4y_5y_6y_8y_9y_{10}.$$

Thus ten main effects and forty-five two-factor interactions may be estimated from a trial testing 128 of the 1024 possible treatment combinations.

(*b*) *Sixteenth replication*

If we are testing twelve or more factors a possible identity relationship is the following:

$$I = y_1y_2y_3y_4y_5 = y_1y_2y_6y_7y_8 = y_3y_4y_5y_6y_7y_8 = y_1y_2y_9y_{10}y_{11}$$
$$= y_3y_4y_5y_9y_{10}y_{11} = y_6y_7y_8y_9y_{10}y_{11} = y_1y_2y_3y_4y_5y_6y_7y_8y_9y_{10}y_{11}$$
$$= y_1y_3y_6y_9y_{12} = y_2y_4y_5y_6y_9y_{12} = y_2y_3y_7y_8y_9y_{12} = y_1y_4y_5y_7y_8y_9y_{12}$$
$$= y_2y_3y_6y_{10}y_{11}y_{12} = y_1y_4y_5y_6y_{10}y_{11}y_{12} = y_1y_3y_7y_8y_{10}y_{11}y_{12} = y_2y_4y_5y_7y_8y_{10}y_{11}y_{12}.$$

In this case twelve main effects and sixty-six two-factor interactions may be estimated from a trial testing 256 of the possible 4096 treatment combinations.

The extent to which these designs will be of practical value depends very much on the existence of a sufficient mass of reasonably homogeneous material to test the large number of treatment combinations without the necessity of dividing the material into smaller batches and using the device of confounding. An experiment involving say 256 different treatment combinations is not large by modern standards. At Rothamsted, for example, an experiment involving 200 distinct treatments on 300 plots has been carried out for some years: this experiment was, however, made possible by utilizing the elimination of the effects of soil heterogeneity by highly complex confounding; the design, in fact, consisted of three 5×5 lattice squares necessitating seventy-five plots, and each of these plots was split into four subplots. The advantages of testing twelve factors, say, at the same time under virtually the same experimental conditions cannot, however, be ignored. Such an experiment should have more value, other things being equal, than two distinct experiments each testing some of the factors. An examination has not been made of the possibilities of reducing block size by confounding for the above two designs, but it is probably necessary to sacrifice a few two-factor interactions.

The relationship between fractional replication and confounding

It is clear that fractional replication and confounding are different aspects of the same process. A 2^n design of 2^p blocks may be described as a 1 in 2^p replicate of a 2^{n+p} design with no subdivision into blocks, by regarding the blocks as a 2^p system in p factors. As an

example, consider the 2^5 design in y_1, y_2, y_3, y_4 and y_5 laid out in four blocks of eight and confounding $y_1y_2y_3$, $y_3y_4y_5$ and $y_1y_2y_4y_5$; superimposing two pseudo-factors b_1 and b_2, the experiment is a quarter-replicate of a 2^7 design in $y_1, y_2, y_3, y_4, y_5, b_1, b_2$. The identity on which the quarter replicate is based is given by the equations

$$b_1 = y_1y_2y_3, \quad b_2 = y_3y_4y_5, \quad b_1b_2 = y_1y_2y_4y_5$$

or the equation

$$I = y_1y_2y_3b_1 = y_3y_4y_5b_2 = y_1y_2y_4y_5b_1b_2.$$

If we examine this equation in the same way as in the previous sections, we find that the design depends on the fact that the aliases of the following type may be ignored:

$$y_1 = y_2y_3b_1 = y_1y_3y_4y_5b_2 = y_2y_4y_5b_1b_2,$$
$$y_1y_2 = y_3b_1 = y_1y_2y_3y_4y_5b_2 = y_4y_5b_1b_2.$$

This example is worth pursuing. The design is frequently used with one replication only, the error being estimated from three-factor and higher-order interactions. We set out below the identity and 31 degrees of freedom together with all their aliases and their usual place in the analysis of variance—blocks (B), treatment (T), or error (E). For convenience of printing we denote the factors tested in the experiment by a, b, c, d, e instead of $y_1y_2y_3y_4y_5$ and the block factors by x and y. Capitals are used for treatment effects thus conforming to present usage.

I	$= ABCX$	$= CDEY$	$= ABDEXY$	
A	$= BCX$	$= ACDEY$	$= BDEXY$	T
B	$= ACX$	$= BCDEY$	$= ADEXY$	T
AB	$= CX$	$= ABCDEY$	$= DEXY$	T
C	$= ABX$	$= DEY$	$= ABCDEXY$	T
AC	$= BX$	$= ADEY$	$= BCDEXY$	T
BC	$= AX$	$= BDEY$	$= ACDEXY$	T
ABC	$= X$	$= ABDEY$	$= CDEXY$	B
D	$= ABCDX$	$= CEY$	$= ABEXY$	T
AD	$= BCDX$	$= ACEY$	$= BEXY$	T
BD	$= ACDX$	$= BCEY$	$= AEXY$	T
ABD	$= CDX$	$= ABCEY$	$= EXY$	E
CD	$= ABDX$	$= EY$	$= ABCEXY$	T
ACD	$= BDX$	$= AEY$	$= BCEXY$	E
BCD	$= ADX$	$= BEY$	$= ACEXY$	E
$ABCD$	$= DX$	$= ABEY$	$= CEXY$	E
E	$= ABCEX$	$= CDY$	$= ABDXY$	T
AE	$= BCEX$	$= ACDY$	$= BDXY$	T
BE	$= ACEX$	$= BCDY$	$= ADXY$	T
ABE	$= CEX$	$= ABCDY$	$= DXY$	E
CE	$= ABEX$	$= DY$	$= ABCDXY$	T
ACE	$= BEX$	$= ADY$	$= BCDXY$	E
BCE	$= AEX$	$= BDY$	$= ACDXY$	E
$ABCE$	$= EX$	$= ABDY$	$= CDXY$	E
DE	$= ABCDEX$	$= CY$	$= ABXY$	T
ADE	$= BCDEX$	$= ACY$	$= BXY$	E
BDE	$= ACDEX$	$= BCY$	$= AXY$	E
$ABDE$	$= CDEX$	$= ABCY$	$= XY$	B
CDE	$= ABDEX$	$= Y$	$= ABCXY$	B
$ACDE$	$= BDEX$	$= AY$	$= BCXY$	E
$BCDE$	$= ADEX$	$= BY$	$= ACXY$	E
$ABCDE$	$= DEX$	$= ABY$	$= CXY$	E

If we take for each linear function of the yields the alias involving the smallest possible number of letters, but remembering that x, y are pseudo-factors, so that X, Y and XY are of

equal importance and therefore XY should be regarded as a main effect and not an interaction, we have the following allocation of contrasts to the three components of the analysis of variance:

Blocks:	X, Y, XY.
Treatments:	A, B, C, D, E. $AB = CX$, $AC = BX$, $BC = AX$, $CD = EY$, $DE = CY$, $CE = DY$. AD, BD, AE, BE.
Error:	AY, BY, DX, EX, AXY, BXY, CXY, DXY, EXY, ACD, BCD, ACE, BCE.

The four three-factor interactions could equally well be regarded as interactions between two-factor interactions and blocks. It would be anticipated that these would be smaller than the interactions of main effects and blocks. The purpose of the present exposition is to give a clear statement of the possible interpretations of the results of an individual experiment. Further remarks on the problem of interpretation are postponed to a later section in the paper.

An example of fractional replication with confounding

A design which has proved of practical utility is the half-replicate of a 2^6 experiment arranged in four blocks of eight plots.

Call the factors $y_1, y_2, y_3, y_4, y_5, y_6$. Then the best confounding is that in which, using full replication, the block differences are all third-order interactions, say

$$y_1y_2y_3y_4, \quad y_3y_4y_5y_6 \quad \text{and} \quad y_1y_2y_5y_6.$$

But it is impossible to keep main effects and interactions clear with this confounding, whatever interaction is equated to the identity.

If we take the confounded interactions to be of the type

$$y_1y_2y_3, \quad y_3y_4y_5, \quad y_1y_2y_4y_5,$$

and the interaction $y_1y_2y_3y_4y_5y_6$ to be unity, then the following interactions are also confounded:

$$y_4y_5y_6, \quad y_1y_2y_6 \quad \text{and} \quad y_3y_6.$$

It will be found by enumeration of the possibilities that one first-order interaction must be sacrificed. All main effects and the other first-order interactions will have high-order aliases.

It is interesting to examine this design in the same way as the 2^5 above for the relations between block-treatment interactions and treatment interactions.

There are, in fact, only thirty-two independent contrasts, and it is simplest to enumerate these by operating on the identity relationship with the thirty-two possibilities for the 2^5 system omitting y_6. As before, we insert block pseudo-factors. For simplicity of printing we use A, B, C, D, E, F for the factors and X, Y for the block factors. Then

$$I = ABCDEF, \quad X = ABC, \quad Y = CDE, \quad XY = ABDE,$$

and combining these into one relationship, we have

$$I = ABCDEF = ABCX = CDEY = ABDEXY = DEFX = ABFY = CFXY.$$

A complete table of the aliases for this design follows:

	$= ABCDEF$	$= ABCX$	$= DEFX$	$= CDEY$	$= ABFY$	$= ABDEXY$	$= CFXY$	
	$= BCDEF$	$= BCX$	$= ADEFX$	$= ACDEY$	$= BFY$	$= BDEXY$	$= ACFXY$	T
	$= ACDEF$	$= ACX$	$= BDEFX$	$= BCDEY$	$= AFY$	$= ADEXY$	$= BCFXY$	T
B	$= CDEF$	$= CX$	$= ABDEFX$	$= ABCDEY$	$= FY$	$= DEXY$	$= ABCFXY$	T
	$= ABDEF$	$= ABX$	$= CDEFX$	$= DEY$	$= ABCFY$	$= ABCDEXY$	$= FXY$	T
C	$= BDEF$	$= BX$	$= ACDEFX$	$= ADEY$	$= BCFY$	$= BCDEXY$	$= AFXY$	T
C	$= ADEF$	$= AX$	$= BCDEFX$	$= BDEY$	$= ACFY$	$= ACDEXY$	$= BFXY$	T
BC	$= DEF$	$= X$	$= ABCDEFX$	$= ABDEY$	$= CFY$	$= CDEXY$	$= ABFXY$	B
	$= ABCEF$	$= ABCDX$	$= EFX$	$= CEY$	$= ABDFY$	$= ABEXY$	$= CDFXY$	T
D	$= BCEF$	$= BCDX$	$= AEFX$	$= ACEY$	$= BDFY$	$= BEXY$	$= ACDFXY$	T
D	$= ACEF$	$= ACDX$	$= BEFX$	$= BCEY$	$= ADFY$	$= AEXY$	$= BCDFXY$	T
BD	$= CEF$	$= CDX$	$= ABEFX$	$= ABCEY$	$= DFY$	$= EXY$	$= ABCDFXY$	E
D	$= ABEF$	$= ABDX$	$= CEFX$	$= EY$	$= ABCDFY$	$= ABCEXY$	$= DFXY$	T
CD	$= BEF$	$= BDX$	$= ACEFX$	$= AEY$	$= BCDFY$	$= BCEXY$	$= ADFXY$	E
CD	$= AEF$	$= ADX$	$= BCEFX$	$= BEY$	$= ACDFY$	$= ACEXY$	$= BDFXY$	E
BCD	$= EF$	$= DX$	$= ABCEFX$	$= ABEY$	$= CDFY$	$= CEXY$	$= ABDFXY$	T
	$= ABCDF$	$= ABCEX$	$= DFX$	$= CDY$	$= ABEFY$	$= ABDXY$	$= CEFXY$	T
E	$= BCDF$	$= BCEX$	$= ADFX$	$= ACDY$	$= BEFY$	$= BDXY$	$= ACEFXY$	T
E	$= ACDF$	$= ACEX$	$= BDFX$	$= BCDY$	$= AEFY$	$= ADXY$	$= BCEFXY$	T
BE	$= CDF$	$= CEX$	$= ABDFX$	$= ABCDY$	$= EFY$	$= DXY$	$= ABCEFXY$	E
E	$= ABDF$	$= ABEX$	$= CDFX$	$= DY$	$= ABCEFY$	$= ABCDXY$	$= EFXY$	T
CE	$= BDF$	$= BEX$	$= ACDFX$	$= ADY$	$= BCEFY$	$= BCDXY$	$= AEFXY$	E
CE	$= ADF$	$= AEX$	$= BCDFX$	$= BDY$	$= ACEFY$	$= ACDXY$	$= BEFXY$	E
BCE	$= DF$	$= EX$	$= ABCDFX$	$= ABDY$	$= CEFY$	$= CDXY$	$= ABEFXY$	T
E	$= ABCF$	$= ABCDEX$	$= FX$	$= CY$	$= ABDEFY$	$= ABXY$	$= CDEFXY$	T
DE	$= BCF$	$= BCDEX$	$= AFX$	$= ACY$	$= BDEFY$	$= BXY$	$= ACDEFXY$	E
DE	$= ACF$	$= ACDEX$	$= BFX$	$= BCY$	$= ADEFY$	$= AXY$	$= BCDEFXY$	E
BDE	$= CF$	$= CDEX$	$= ABFX$	$= ABCY$	$= DEFY$	$= XY$	$= ABCDEFXY$	B
DE	$= ABF$	$= ABDEX$	$= CFX$	$= Y$	$= ABCDEFY$	$= ABCXY$	$= DEFXY$	B
CDE	$= BF$	$= BDEX$	$= ACFX$	$= AY$	$= BCDEFY$	$= BCXY$	$= ADEFXY$	T
CDE	$= AF$	$= ADEX$	$= BCFX$	$= BY$	$= ACDEFY$	$= ACXY$	$= BDEFXY$	T
$ABCDE$	$= F$	$= DEX$	$= ABCFX$	$= ABY$	$= CDEFY$	$= CXY$	$= ABDEFXY$	T

The partition of the degrees of freedom in the analysis of variance which would generally be made is the following:

		D.F.
Blocks		3
Treatments:	Main effects	6
	Interactions	14
Error		8
		31

The table of aliases is condensed below by the omission of all aliases involving more than two factors—counting, as before, XY as a single factor as well as X and Y.

Effects A, B, D, E have aliases of at least three letters, but $C = FXY$ and $F = CXY$.

Effects AD, BD, AE, BE have aliases of at least three letters, but

$$AB = CX = FY, \quad AC = BX, \quad BC = AX, \quad CD = EY, \quad EF = DX,$$

$$CE = DY, \quad DE = FX = CY, \quad DF = EX, \quad BF = AY, \quad AF = BY.$$

In an experiment in which block-treatment interactions cannot be assumed to be negligible in relation to the effects it is desired to estimate, the interpretation of most two-factor interactions is difficult if not impossible. The following identities of practical interest exist for the terms which would be used to estimate the error: ACD, BCD, ACE, BCE have aliases of three letters and are either three-factor interactions or interactions between blocks and two-factor interactions, but $ABD = EXY$, $ABE = DXY$, $ADE = BXY$, and $BDE = AXY$.

This design is very similar in result to the fully replicated but confounded 2^5 design described above.

Fractional replication in the 3^n system

Here we have to consider treatment effects assessed from powers of one-third of a complete replicate. Only those treatment combinations represented by points of the lattice lying on the hyperplane $\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_n y_n = 0$, or 1, or 2 (mod 3)

will be included in a one-third replicate.

A particular treatment effect is given by the differences between the means of those plots represented by points on the following three planes:

$$\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 0 \pmod 3,$$
$$\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 1 \pmod 3,$$
$$\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 2 \pmod 3.$$

It is obvious that the points lying on the first plane will also lie on the planes

$$(\beta_1 + \lambda\alpha_1) y_1 + (\beta_2 + \lambda\alpha_2) y_2 + \ldots + (\beta_n + \lambda\alpha_n) y_n = 0 \pmod 3, \quad \text{for } \lambda = 1 \text{ and } 2;$$

the points on the other two planes will lie on these planes with 1 and 2 respectively on the right-hand side of the equation.

The aliases of each pair of degrees of freedom are therefore obtained by multiplication of its symbol by $y_1^{\alpha_1} y_2^{\alpha_2} \ldots y_n^{\alpha_n}$,

and by its square.

As an example, suppose a third replicate of a 3^3 design is based on the inclusion only of those treatment combinations represented by the symbol $y_1 y_2 y_3 (y_1 + y_2 + y_3 = 0$ say), then the aliases are exemplified by the relationship $y_1 = y_1 y_1^2 y_3^2 = y_2 y_3$.

The confounding of one replicate of a 3^3 experiment in three blocks of nine plots

A frequently used design is the 3^3 in three blocks of nine plots, testing all combinations of three factors each at three levels. This design is formally a one-third replicate of a 3^4 design. Suppose the factors are y_1, y_2, and y_3 and let blocks be denoted by the pseudo-factor b; a three-factor interaction of y_1, y_2, and y_3, say $y_1 y_2 y_3$, is usually confounded in order to keep main effects and first-order interactions free of block effects.

Then $b = y_1 y_2 y_3$ or $I = y_1 y_2 y_3 b^2$, since $b^3 = 1$.

As in the case of the 2^5 design, we work out the aliases of each pair of degrees of freedom: each pair of degrees of freedom will in this case have two aliases:

$$\begin{aligned}
y_1 &= y_1 y_2^2 y_3^2 b &&= y_2 y_3 b^2 & \qquad y_2 y_3 &= y_1 y_2^2 y_3^2 b^2 &&= y_1 b^2 \\
y_2 &= y_1 y_2^2 y_3 b^2 &&= y_1 y_3 b^2 & \qquad y_2 y_3^2 &= y_1 y_2^2 b^2 &&= y_1 y_3^2 b^2 \\
y_3 &= y_1 y_2 y_3^2 b^2 &&= y_1 y_2 b^2 & \qquad y_1 y_2 y_3 &= y_1 y_2 y_3 b &&= b^2 \\
y_1 y_2 &= y_1 y_2 y_3^2 b &&= y_3 b^2 & \qquad y_1 y_2 y_3^2 &= y_1 y_2 b &&= y_3 b \\
y_1 y_3^2 &= y_1 y_3^2 b &&= y_2 y_3^2 b & \qquad y_1 y_2^2 y_3 &= y_1 y_3 b &&= y_2 b \\
y_1 y_3 &= y_1 y_2^2 y_3 b &&= y_2 b^2 & \qquad y_1 y_2^2 y_3^2 &= y_1 b &&= y_2 y_3 b \\
y_1 y_3^2 &= y_1 y_2^2 b &&= y_2 y_3^2 b^2 &&&&
\end{aligned}$$

Here again the identities could result in difficulty in interpretation—as of course could have been predicted from the examination of the possible arrangements in blocks of nine of the 3^4 design. The main effects may be regarded as clear, and three of the first-order interactions. The remaining two-factor interactions could be ascribed to differential effects of the factors on the three blocks. The three-factor interactions which are not confounded with blocks are also ascribable to interactions of main effects and blocks and may therefore be used to form an estimate of the error of these effects.

GENERAL REMARKS ON CONFOUNDING

The device of confounding is used almost without exception in agricultural experiments in order to reduce the block size to twelve or less plots. As the above results indicate there are two aspects which then need careful consideration, (*a*) the estimation of interactions, and (*b*) the estimation of the experimental error.

The main purpose of the factorial design is the estimation of main effects and interactions between pairs of factors and thence of the effect of any one factor in the presence and absence of each of the other factors. It is clear that when it is necessary to remove soil heterogeneity by confounding, the interpretation of a small experiment involving a few factors may be exceedingly difficult because of the possibility of block-treatment interactions. It is possible to use the rule that a large contrast should be regarded as the interaction between whichever pair of main effects is the larger, but this rule will break down in some cases when, for example, the contrast has two aliases AB and CD, and effects A and C are large and B and D small. In the case of a series of experiments, a device which might be helpful is the use of permutations of the possible identity relationships, one at each centre. The modern emphasis in agricultural experimentation is on series of experiments at various places and in several years, rather than on individual experiments. Interactions of pairs of factors will be estimated correctly from a large series of experiments if treatments are assigned at random to blocks.

The evaluation of two-factor interactions for individual experiments depends on the assumption that block-treatment interactions are small compared with the experimental error. Yates (1935) examined several experiments for the existence of such interactions and found no evidence of them. Since that time a large number of experimental results which can be used to provide information on the question have been accumulated, and an investigation of these has indicated that block-treatment interactions are negligible and may be ignored (Kempthorne, 1947).

With regard to the estimation of error, in so far as tests of significance are of interest, it can be said that the analysis of variance does provide a test of significance of the hypothesis that the treatments have an overall effect different from zero. In agricultural experimentation, the term error is used to denote block-treatment interactions. Thus in the simple randomized block experiment, it is possible to evaluate the difference between two treatments from each block, and it is the variability of this difference from block to block which is regarded as the error. In general, as there are usually few blocks, and the error of each comparison would be determined with poor accuracy, the errors of all the possible independent comparisons are pooled to give a common estimate. If the treatments were duplicated at random

within each block, the analysis would be of the form (r being the number of blocks and t of treatments):

	D.F.
Blocks	$r-1$
Treatments	$t-1$
Treatments by blocks	$(r-1)(t-1)$
Within blocks	rt
	$2rt-1$

The component 'within blocks' could more accurately be described as experimental error, but would not be used to evaluate the errors of treatment effects, since the experimenter is interested in the constancy of treatment effects from block to block. There is therefore little point in actually carrying out such an experiment. In a factorial experiment with replication, the components which could be evaluated consist of replicates, effects and low-order interactions, high-order interactions, and interactions of treatments and replicates. On the assumption that the sum of squares for interaction of treatments and replicates is homogeneous, the mean square for high-order interactions will include the mean square for treatments × replicates plus a component of variance due to high-order interactions. When only one replication is used, it is assumed that the component of variance due to high-order interactions is small, and that the high-order interactions mean square can be regarded as an estimate of error. It is important to bear in mind that an individual agricultural experiment can give information only for a particular set of experimental conditions and that it is known from experience that place to place and year to year variability is considerable. It would therefore be uneconomical to utilize available resources to determine effects and their errors at a few particular places very accurately, but preferable to sacrifice replication at each place in order to have information over a large range of experimental conditions.

Mixed systems

It is not proposed to examine mixed systems of the type $p^m q^n$, where p and q are primes, in the present paper. It is clear, however, that the possibilities of complete confounding and fractional replication are very limited. A p'th replicate must obviously include p^{m-1} combinations of the m factors combined with all the q^n combinations of the n factors. For the examination of treatment aliases the system may be regarded as the product of the two separate systems. Thus if $p = 3$, $m = 2$, $q = 2$, $n = 3$ and the factors are $y_1 y_2 y_3' y_4' y_5'$, then a half replicate would be obtained by putting $I = y_3' y_4' y_5'$. The aliases which result are exemplified by the following:

$$y_1 = y_1 y_3' y_4' y_5', \qquad y_1 y_2 y_3' = y_1 y_2 y_4' y_5',$$
$$y_1 y_2 = y_1 y_2 y_3' y_4' y_5', \qquad y_3' = y_4' y_5'.$$

Such designs with fractional replication or complete confounding are therefore useful only when the corresponding designs for the two separate systems are feasible.

Comments on 'the design of optimum multifactorial experiments'

In a paper entitled 'The Design of Optimum Multifactorial Experiments', Plackett & Burman (1946) put forward designs more specifically for physical and industrial research, which are of interest from the point of view of fractional replication. In order to estimate the

effect of varying nine components, of an assembly, each component having two possible values, a nominal (−) and an extreme (+), they put forward the following design which requires the testing of sixteen assemblies:

	Components								
	1	2	3	4	5	6	7	8	9
Assembly 1	+	−	−	−	+	−	−	+	+
2	+	+	−	−	−	+	−	−	+
3	+	+	+	−	−	−	+	−	−
4	+	+	+	+	−	−	−	+	−
5	−	+	+	+	+	−	−	−	+
6	+	−	+	+	+	+	−	−	−
7	−	+	−	+	+	+	+	−	−
8	+	−	+	−	+	+	+	+	−
9	+	+	−	+	−	+	+	+	+
10	−	+	+	−	+	−	+	+	+
11	−	−	+	+	−	+	−	+	+
12	+	−	−	+	+	−	+	−	+
13	−	+	−	−	+	+	−	+	−
14	−	−	+	−	−	+	+	−	+
15	−	−	−	+	−	−	+	+	−
16	−	−	−	−	−	−	−	−	−

Yates put forward a similar design in his 1935 paper for the weighing of a number of small articles on a balance which required a zero correction, as an example of the estimation of the effects of independent factors. In his case there was a close formal analogy to the 2^n factorial system, and it will now be shown that Plackett & Burman's design given above is a high-order fractional design of the type discussed in the present paper.

Denoting the nominal values by unity and the extreme values of the nine components by $a, b, c, d, e, f, g, h, k$ in order, the treatment combinations represented are l, $aehi$, $abfi$, $abcg$, $abcdh$, $bcdei$, $acdef$, $bdefg$, $acefgh$, $abdfghi$, $bceghi$, $cdfhi$, $adegi$, $befh$, $cfgi$, dgh. It is found merely by one-by-one examination of the three-factor interactions that all the above sets of treatment combinations occur with the same sign in the following:

$$ABE, \quad ACK, \quad BCF, \quad CDG, \quad DEH.$$

The same will be true for all the members of the Abelian group of which the above five interactions are generators. The identity relationship is therefore:

$$\begin{array}{llllllll} I & = ABE & = ACK & = BCEK & = BCF & = ACEF & = ABFK & = EFK \\ = CDG & = ABCDEG & = ADGK & = BDEGK & = BDFG & = ADEFG & = ABCDFGK & = CDEFGK \\ = DEH & = ABDH & = ACDEHK & = BCDHK & = BCDEGH & = ACDFH & = ABDEFHK & = DFHK \\ = CEGH & = ABCGH & = AEGHK & = BGHK & = BEFGH & = AFGH & = ABCEFGHK & = CFGHK \end{array}$$

The identities of interest to the experimenter are the following:

$$I = ABE = ACK = BCF = EFK = CDG = DEH;$$

from these we derive the following aliases for main effects:

$$\begin{array}{ll} A = BE = CK, & F = BC = EK, \\ B = AE = CF, & G = CD, \\ C = AK = BF = DG, & H = DE, \\ D = CG = EH, & K = AC = EF. \\ E = AB = FK = DH, & \end{array}$$

In all cases, the contrasts estimating main effects are minus the contrasts estimating interactions. If, for example, the interaction of B and E is negative, and A has no effect, the

conclusion drawn by the experimenter will be that A has a positive effect. It is possible but rather difficult to imagine physical systems in which effects will not interact, and interpretation of the results of experiments based on this design may often be impossible. With nine factors, it appears from the present work that the minimum number of combinations which should be tested is 128, that is one-quarter of a replication, though it is possible that by making less stringent assumptions about two-factor interactions, one-eighth of a replication might give intelligible results. A possible instance in which it might be feasible to use the designs discussed is when it is expected that only one or two of the factors have an effect, and the problem is to determine as quickly as possible which of the nine factors are responsible. An example in which a high-order fractional design was used in such circumstances with good results has been described by Tippett (1936). A detailed examination of all the designs put forward by Plackett & Burman will not be undertaken, but the lines on which such an examination would proceed and the broad conclusions which would emerge are obvious from the above examination of one of their simpler designs.

Conclusions

A method of examining fractional replication and confounding for some types of factorial experiments is described. The formal equivalence between the two is indicated and the implications of this equivalence discussed. Further progress will follow on group theory lines and this is being examined, together with the possibility of fractional replication when the fraction is greater than unity. The possibilities are explored of the estimation of main effects and two-factor interactions of many factors by testing only a small proportion of the possible treatment combinations. An examination on these lines is made of designs proposed by Plackett & Burman.

REFERENCES

Finney, D. J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen., Lond.*, **12**, 291–301.

Fisher, R. A. (1942). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen., Lond.*, **11**, 341–53.

Kempthorne, O. (1947). A note on differential responses in blocks. *J. Agric. Sci.* **37**, 245–48.

Plackett, R. L. & Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–25.

Tippett, L. H. C. (1936). *Applications of Statistical Methods to the Control of Quality in Industrial Production*. Manchester Statist. Soc.

Yates, F. (1935). Complex experiments. *J. R. Statist. Soc. Suppl.* **2**, 181–223.

Yates, F. (1937). The design and analysis of factorial experiments. *Tech. Commun. Imp. Bur. Soil Sci.*, no. 35.

A COMPARISON OF STRATIFIED WITH UNRESTRICTED RANDOM SAMPLING FROM A FINITE POPULATION*

By P. ARMITAGE, B.A.

1. Introduction

1·1. We are concerned in this paper with the problem of estimating the mean value μ of a variable x in a population, by taking a sample which is in some way representative of the population. It has been realized since Bowley's paper (1926), and more particularly since Neyman's more comprehensive survey (1934), that a certain degree of precision in the estimate can often be obtained more economically by stratified random sampling (usually referred to merely as *stratified sampling*) than by unrestricted random sampling (usually called merely *random sampling*). In the stratified method, the population is divided into several strata, the sample size divided in some prearranged way among the strata, and sampling performed at random from each stratum. In unrestricted random sampling, a random selection is made from the whole population, and the method may be regarded as a particular case of stratification, where the number of strata is one.

Some text-books deal briefly with stratified sampling. Wilks (1943) considers only infinite populations, and denotes by representative sampling what we should call a particular type of stratified sampling (see § 1·2). The subject is treated by Kendall (1946, pp. 249–52), but he makes no comparison with unrestricted random sampling. We shall begin by introducing several well-known results which will be needed later.

1·2. The summation sign Σ will be used throughout for $\sum_{l=1}^{r}$, and $\sum_{k}$ for $\sum_{k=1}^{r}$. In general, Σ is used for a single summation, $\sum_{k}\Sigma$ for a double summation, and the suffix k where no summation is involved.

We shall consider the following position: A population π of size N is subdivided into r strata, π_k, of size N_k ($\Sigma N_l = N$). The variable x is distributed so that the mean and variance (divisor N_k) within π_k are respectively μ_k, σ_k^2. It is required to estimate $\mu = \Sigma N_l \mu_l / N$, the grand mean.

Suppose a given sample size, n, is divided so that n_k items are sampled at random from π_k ($\Sigma n_l = n$). We may denote the jth observation from the kth sample by x_{kj} ($j = 1, 2, \ldots, n_k$), and the mean and variance of the kth sample by $\bar{x}_k$ and s_k^2, which are known to be unbiased estimates of μ_k and $\dfrac{(n_k - 1) N_k \sigma_k^2}{n_k (N_k - 1)}$, respectively (see, for example, Kendall, 1943, p. 284). It seems intuitively obvious to take as our estimate of μ,

$$m = \Sigma N_l \bar{x}_l / N, \qquad \text{(i)}$$

which is clearly unbiased. This is, however, not the only unbiased estimate which is a linear function of the x_{kj}. For instance, $\Sigma N_l x_{l1} / N$ also satisfies the conditions. Neyman (1934) has shown that, for fixed values of n_k, the estimate given by (i) is the best linear unbiased estimate of μ, in the sense that its sampling variance is less than that of any other linear unbiased estimate.

* Communication from the National Physical Laboratory.

The question now arises: given a sample size n, how shall we choose the n_k so as to minimize var (m), where m is given by (i)? Bowley had not considered 'best' estimates, and he suggested that n_k should be proportional to N_k, i.e.

$$n_k = \frac{nN_k}{N}. \tag{ii}$$

Neyman (1934) showed, by the method given in § 2, that the values of n_k which minimize var (m) are

$$n_k = \frac{nN_k\sigma_k\sqrt{[N_k/(N_k-1)]}}{\sum N_l\sigma_l\sqrt{[N_l/(N_l-1)]}}$$

$$= \frac{nN_k\sigma'_k}{\sum N_l\sigma'_l}, \tag{iii}$$

where $\sigma'_k = \sigma_k\sqrt{[N_k/(N_k-1)]}$.

We shall refer to these two methods of defining the n_k, by (ii) and (iii) respectively, as *proportionate sampling*, and *optimum stratified sampling*, denoting by m_p and m_o the estimates of μ obtained from (i) by the two methods, and by $\bar{x}$ the estimate of μ given by the mean of an unrestricted random sample of n from the whole population π.

The optimum stratified method thus requires a knowledge of the σ_k. In practice, we should never know the σ_k exactly, unless the population had been subjected to exhaustive sampling, in which case μ would be known exactly. Sukhatme (1935) has shown that, at any rate for large N_k, if the σ_k^2 are estimated from a preliminary sample, and the n_k defined by using these estimates in (iii), there is a high probability that var $(m_o) <$ var (m_p).* The efficiency of this method will of course depend on the size of the preliminary sample, and Sukhatme's investigation only dealt with one value of this (15 from each stratum). In some cases we should be able to form a fairly good estimate of the σ_k from past experience, and there would be no need for a preliminary sample.

Another interesting comparison which has not been extensively investigated is that between optimum stratified sampling and unrestricted random sampling. Wilks (1943) deals with this for infinite populations, and obtains (pp. 88, 89) the result (in our notation),

$$\text{var}\,(m_o) \leqslant \text{var}\,(m_p) \leqslant \text{var}\,(\bar{x}), \tag{iv}$$

the first equality holding only when all the σ_k are equal, and the second only when all the μ_k are equal. (Our N_k/N are replaced by p_k, where p_k is the probability that x, when drawn at random from π, is a member of π_k, so that, for instance, (iii) becomes

$$n_k = \frac{np_k\sigma_k}{\sum p_l\sigma_l}.$$

Representative sampling as defined by Wilks is what we should call proportionate sampling.) We shall show in § 2 that for finite populations, while the relation

$$\text{var}\,(m_o) \leqslant \text{var}\,(m_p) \tag{v}$$

is always true, the equality holding only when all the σ'_k are equal, it is not necessarily true that

$$\text{var}\,(m_p) \leqslant \text{var}\,(\bar{x}), \tag{vi}$$

* No confusion need arise from the fact that the symbol m_o and the term *optimum* are still used when estimates of the σ_k are used in (iii).

and in fact in the limiting case when all the μ_k are equal, it is true that

$$\text{var}(m_p) > \text{var}(\bar{x}), \tag{via}$$

so that if the σ'_k are also equal

$$\text{var}(m_o) > \text{var}(\bar{x}); \tag{vib}$$

i.e. random sampling gives a more accurate estimate of the mean than any stratified sampling. We shall see, however, that in almost all practical cases (iv) is true.

2. Derivation of formulae

2·1. *Results* (iii) *and* (v). Using the notation of § 1·2, we have the standard result that

$$\text{var}(\bar{x}_k) = \frac{\sigma_k^2}{n_k}\left(\frac{N_k - n_k}{N_k - 1}\right) \quad \text{(see e.g. Wilks, p. 86).} \tag{vii}$$

Therefore from (i),

$$\text{var}(m) = \sum \frac{N_l^2 \sigma_l^2}{N^2 n_l}\left(\frac{N_l - n_l}{N_l - 1}\right)$$

$$= \sum \frac{N_l \sigma_l'^2}{N^2 n_l}(N_l - n_l). \tag{viii}$$

The result (iii) may be obtained quite easily by finding the values of the n_l which minimize (viii) subject to the condition $\sum n_l = n$, using the method of Lagrange multipliers. Then, substituting (ii) and (iii) in (viii), and applying Schwarz's inequality, we have (v). The following method is due to Neyman.

It may be verified from (viii) that

$$\text{var}(m) = \frac{N-n}{N^2 n}\sum N_l \sigma_l'^2 + \frac{1}{N^2}\sum_k n_k\left(\frac{N_k \sigma'_k}{n_k} - \frac{\sum N_l \sigma'_l}{n}\right)^2 - \frac{1}{Nn}\sum_k N_k\left(\sigma'_k - \frac{\sum N_l \sigma'_l}{N}\right)^2. \tag{ix}$$

If we denote the three terms of (ix) by A, B and C, so that

$$\text{var}(m) = A + B - C,$$

it will be seen that A and C are independent of n_k and, since B is non-negative, it follows that the values of n_k which minimize var (m) must minimize B. Now $B = 0$ if and only if

$$n_k = \frac{n N_k \sigma'_k}{\sum N_l \sigma'_l},$$

which is (iii). For these values of n_k, $m = m_o$, and

$$\text{var}(m_o) = A - C. \tag{x}$$

If we define n_k by (ii), so that $m = m_p$, we see from (ix) that $B = C$, so that

$$\text{var}(m_p) = A. \tag{xi}$$

From (x) and (xi), we obtain (v), the equality holding only when $C = 0$, which is true only when $\sigma'_k - \dfrac{\sum N_l \sigma'_l}{N} = 0$ for all k, i.e. when the σ'_k are all equal.

2·2. *Unrestricted random sampling.* The variance of a random observation x from π is

$$\sigma^2 = \frac{\sum N_l \sigma_l^2}{N} + S,$$

where S is the weighted sum of squares of the μ_l, i.e. $S = \frac{\sum N_l(\mu_l - \mu)^2}{N}$. From (vii),

$$\begin{aligned}\text{var}(\bar{x}) &= \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\\ &= \frac{N-n}{nN(N-1)}\sum N_l\sigma_l^2 + \frac{N-n}{n(N-1)}S\\ &= \frac{N-n}{nN(N-1)}\sum(N_l-1)\sigma_l^2 + \frac{N-n}{n(N-1)}S.\end{aligned}$$

From (xi),
$$\text{var}(m_p) = \frac{N-n}{N^2n}\sum N_l\sigma_l^2.$$

Denoting $\frac{\sum(N_l-1)\sigma_l^2}{N-1}$ by H, and $\frac{\sum N_l\sigma_l^2}{N}$ by K, we have

$$\left.\begin{aligned}\text{var}(\bar{x}) &= \frac{N-n}{Nn}H + \frac{N-n}{(N-1)n}S\\ \text{and}\qquad \text{var}(m_p) &= \frac{N-n}{Nn}K.\end{aligned}\right\} \qquad \text{(xii)}$$

Now
$$H - K = \frac{\sum N_l\sigma_l^2 - \sum N_l\sum\sigma_l^2}{N(N-1)} < 0,$$

and if we regard each N_l as being of the same order, $O(N)$, then $H-K$ is $O(N^{-1})$, which means that when all the μ_k are equal, $S = 0$, and so

$$\text{var}(\bar{x}) < \text{var}(m_p),$$

which is (via); but as $N \to \infty$, $\quad \text{var}(\bar{x}) \sim \text{var}(m_p) + S/n,$ (xiii)

giving Wilks's result (p. 88) that for infinite populations (vi) is true, the equality holding only when all the μ_k are equal.

From (v) and (via) it follows that for finite populations, when all the μ_k are equal and all the σ_k' are equal, (vib) is true, i.e. in this case unrestricted random sampling is actually better than any stratified random sampling with the same sample size.

3. General comparison

3·1. From (ix), (x) and (xii),

$$\phi \equiv \text{var}(\bar{x}) - \text{var}(m_c) - \frac{N-n}{(N-1)n}S = \frac{1}{N^2(N-1)n}(P-Q-R), \qquad \text{(xiv)}$$

where
$$\begin{aligned}P &= N^2\sum N_l\sigma_l^2 - N(\sum N_l\sigma_l^2)^2 \geqslant 0 \quad \text{(equality if all } \sigma_l' \text{ are equal).}\\ Q &= n(\sum N_l\sigma_l^2 - N\sum\sigma_l^2) < 0,\\ R &= N^2\sum\sigma_l^2 - (\sum N_l\sigma_l^2)^2 > 0.\end{aligned}$$

As $N \to \infty$, P, Q and R are respectively $O(N^3)$, $O(N)$ and $O(N^2)$, and so we have the result that for infinite populations $\phi \geqslant 0$, which with (xiii) is easily seen to be equivalent to Wilks's result (iv).

In the finite case, however, by suitable choice of the σ_l' and n we can make ϕ either positive or negative. For instance, if the σ_l' are all equal and n is sufficiently small, R predominates in (xiv), and $\phi < 0$. As n increases to N, ϕ increases to 0. (By considering Q and R, it is not

obvious that $\phi \to 0$ in this case, but it must be remembered that (xiv) is only true if the n_k are given by (iii), and this becomes impossible as n approaches N. This will be remarked upon below.) If the σ_i' are sufficiently unequal, P will predominate and $\phi > 0$. In this case the factor $\frac{(P-R)}{N^2(N-1)n}$ in (xiv) will be positive, and ϕ will decrease as n increases.

The situations, then, in which (vi*b*) is likely to be true (provided that the n_k are really given by (iii)) are when the μ_k are nearly equal, and when N is small or the σ_k' are nearly equal. We shall consider some examples in § 4.

3·2. In applying the procedure of stratification, we shall make two departures from the theory outlined above which will tend to nullify the advantages of the stratified method. The first is that, as was pointed out in § 1·2, we shall never know the σ_k exactly, and the degree to which our estimates from which the n_k were obtained are accurate depends on the circumstances. It seems quite likely that Sukhatme's result will be fairly well applicable to finite populations, but there is an opportunity for research on this point.

The second respect in which we depart from theory lies in the fact that, even if the σ_k are exactly known, the n_k that we choose can never be exactly as given by (iii); first because they must be integers, which makes a considerable difference when n is small (the size of the smallest stratified sample from which an unbiased estimate of μ can be made is clearly r); and secondly, n_k cannot take values greater than N_k. In this latter case, if the values of, say, s of the n_k, as given by (iii), are greater than the corresponding N_k, we should let $n_k = N_k$ for these s strata, and then set the other $(r-s)$ values of n_k proportional to the corresponding $N_k\sigma_k'$. This will clearly decrease $\mathrm{var}(\bar{x}) - \mathrm{var}(m_o)$ as given by (xiv). For example, when $n = N$, we have

$$\mathrm{var}(\bar{x}) = \mathrm{var}(m_o) = \frac{(N-n)}{(N-1)n} S = 0,$$

but the right-hand side of (xiv)

$$= \frac{1}{N^3}\{N\Sigma N_i\sigma_i'^2 - (\Sigma N_i\sigma_i')^2\} \geqslant 0$$

(equality holding if all the σ_i' are equal). In fact both these limitations will decrease the theoretical advantage (if any) of stratified over random sampling, and we must take them into account in assessing the relative merits of the two methods.

4. EXAMPLES

In the four examples illustrated by Figs. 1–4, $\mathrm{var}(m_o)$ and $\mathrm{var}(\bar{x})$ have been calculated for different stratified populations, and $\psi = \log_{10}\{\mathrm{var}(\bar{x})/\mathrm{var}(m_o)\}$ plotted against $c = n/N$, so that $\psi < 0$ if $\mathrm{var}(\bar{x}) < \mathrm{var}(m_o)$. In each figure the different curves represent populations with the same σ_k, with the N_k in the same proportions but with different magnitudes, and with the μ_k equal, so that $S = 0$.

Example 1.	$\sigma_k = 2, 3, 4,$	$N_k \propto 6, 4, 3$	$(N = 65, 26, 13)$.
Example 2.	$\sigma_k = 4, 5, 6,$	$N_k \propto 6, 5, 4$	$(N = 120, 60, 30, 15)$.
Example 3.	$\sigma_k = 4, 5, 6,$	$N_k \propto 3, 11, 4$	$(N = 126, 54, 36, 18)$.
Example 4.	$\sigma_k = 1, 1, 2, 3, 4,$	$N_k \propto 5, 5, 1, 2, 3$	$(N = 128, 32)$.

The first thing to be noticed about the graphs is that in each one ψ increases, generally speaking, as n increases. Further, in any one example the range of c for which $\psi < 0$ increases

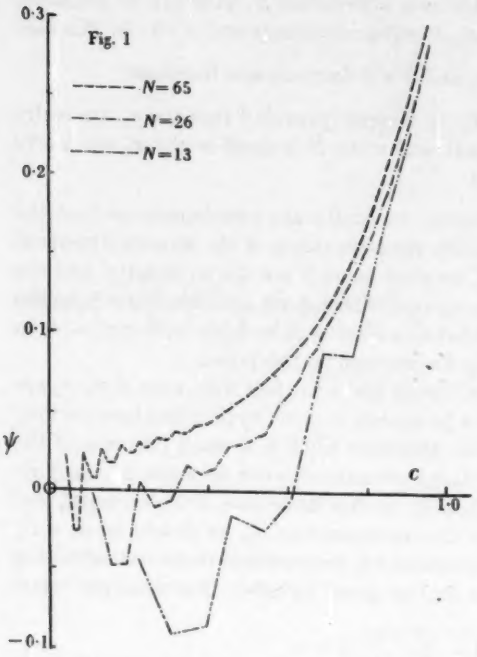


Fig. 1
N = 65
N = 26
N = 13
0·3
0·2
0·1
0
−0·1
ψ
c
1·0

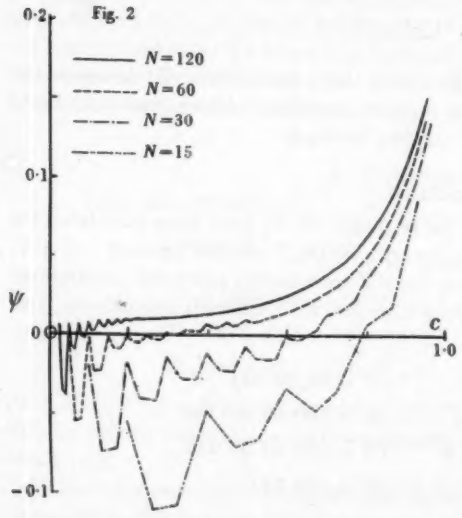


Fig. 2
N = 120
N = 60
N = 30
N = 15
0·2
0·1
0
−0·1
ψ
c
1·0

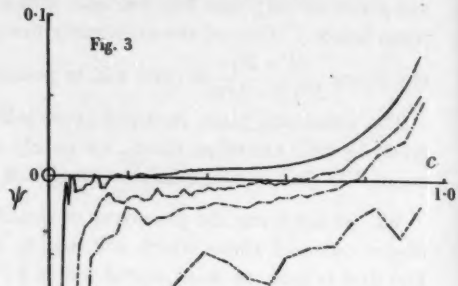


Fig. 3
N = 126
N = 54
N = 36
N = 18
0·1
0
−0·1
−0·2
ψ
c
1·0

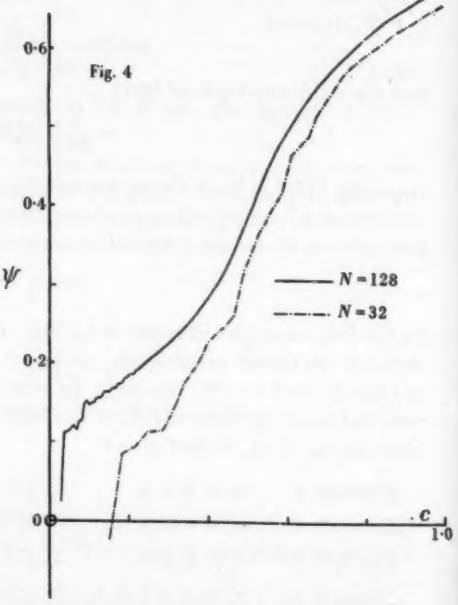


Fig. 4
N = 128
N = 32
0·6
0·4
0·2
0
ψ
c
1·0

as N decreases; and in this sense we can say that for small samples of proportionate size from a stratified population, the advantage (if any) of the stratified method decreases as N decreases.

Secondly, the curves are not smooth. The reason for this is clear. In the optimum stratified method the n_k are to be chosen approximately proportional to $N_k\sigma_k$ (a second approximation is $(N_k+\frac{1}{2})\sigma_k$). In Example 1, the $N_k\sigma_k$ are all equal, and it follows that the n_k should be nearly equal. If $n \equiv 0 \pmod 3$ this can be done, but for $n \equiv 1, 2 \pmod 3$, var (m_o) takes values greater than it would if fractional n_k were allowed. This produces a rise in the curve of ψ for $n \equiv 0 \pmod 3$, which gradually disappears as n increases since the effect is much greater for small n. The same 'period' is noticeable in Fig. 2, but in Figs. 3 and 4, where the main 'periods' are respectively 15 and 30, the effect is smaller.

We saw in § 3·1 that, broadly speaking, the advantage of the stratified method decreases as the σ_k tend to equality. This is illustrated by comparing Examples 1 and 2. In each of these the $N_k\sigma_k$ are equal, but in Example 2 the σ_k are proportionally more nearly equal. Comparing curves for about the same N ($N = 65, 26, 13$ in Fig. 1 with $N = 60, 30, 15$ in Fig. 2), we see that in Fig. 2 the range of values of c for which $\psi < 0$ is greater than in Example 1.

Fig. 3 has the same σ_k as Fig. 2, but the $N_k\sigma_k$, and therefore the n_k, are different. The curves are similar to those of Fig. 2, but the stratified method is still less advantageous (especially for small values of c).

Example 4 has five instead of three strata, and there is quite large variation between the σ_k and between the $N_k\sigma_k$. There is no doubt here that $\psi > 0$, the only exception being for $N = 32$, $n = 5$, where $\psi = -0{\cdot}02$.

These examples may be said to give the maximum advantage to the stratified method, in the sense that the calculated values of var (m_o) depend on the best method of choosing the n_k. If the σ_k are not sufficiently well known to enable the best values of n_k to be used, then we shall get a larger value of var (m_o). It must be remembered, however, that in all these examples we assumed that there was no variation between the μ_k, a situation which would be very unlikely to occur in practice. Now it is clear from (xii) that if the same N_k and σ_k are considered as in one of the above examples, but the μ_k are now unequal, the effect is to increase the value of var $(\bar{x})$ by $(N-n)S/(N-1)n$, where $S = \sum N_i(\mu_i-\mu)^2/N$; so, in any example where $\psi < 0$ for some particular values of N and n, we can reverse the direction of the inequality by choosing a sufficiently large value of S, say

$$S_0 = [\text{var}\,(m_o) - \text{var}\,(\bar{x})]\,(N-1)\,n/(N-n).$$

In comparing different values of S_0 for different examples, it must be remembered that the order of magnitude of S_0 depends on the σ_k and a suitable measure of comparison will be S_0/σ_0^2, where σ_0^2 is the pooled variance within strata $= \sum N_i\sigma_i^2/N$.

In Example 1, the largest value of S_0 is for $N = 13$, $n = 4$. Here var $(m_o) = 1{\cdot}9172$, var $(\bar{x}) = 1{\cdot}5577$, and $S_0 = 1{\cdot}917 = 0{\cdot}231\sigma_0^2$. (If $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 3{\cdot}5$, then $S = 2{\cdot}066$.)

In Example 2, the largest value of S_0 is for $N = 15$, $n = 4$. Here var $(m_o) = 24{\cdot}647$, var $(\bar{x}) = 19{\cdot}119$, and $S_0 = 28{\cdot}14 = 0{\cdot}289\sigma_0^2$. (If $\mu_1 = 0$, $\mu_2 = 7$, $\mu_3 = 13$, then $S = 28{\cdot}51$.)

In Example 3, the largest value of S_0 is for $N = 18$, $n = 3$. Here var $(m_o) = 46{\cdot}235$, var $(\bar{x}) = 30{\cdot}523$, and $S_0 = 53{\cdot}42 = 0{\cdot}515\sigma_0^2$. (If $\mu_1 = 0$, $\mu_2 = 8$, $\mu_3 = 17$, then $S = 57{\cdot}5$.)

In Example 4, the largest value of S_0 is for $N = 32$, $n = 5$ (the only occasion in this example where $\psi < 0$). Here var $(m_o) = 0{\cdot}91406$, var $(\bar{x}) = 0{\cdot}87097$, and $S_0 = 0{\cdot}2474 = 0{\cdot}049\sigma_0^2$. (If $\mu_1 = \mu_2 = 0$ and $\mu_3 = \mu_4 = \mu_5 = 1$, then $S = 0{\cdot}285$.)

5. Conclusions

We have seen in § 3 that optimum stratified sampling may give a less accurate estimate of μ than unrestricted random sampling when the μ_k are nearly equal, and when N is small or the σ'_k are nearly equal. The examples of § 4 bear out these conclusions and show that the effect is greatest for small n, Fig. 3 providing an additional suggestion that if the products $N_k\sigma_k$ are widely different the advantage of the stratified method tends to be nullified. In practice, we should probably only apply stratified sampling if we knew that the strata were sufficiently distinct to ensure considerable variation between either the μ_k or the σ_k. In the first case, if nothing much was known about the σ_k and a preliminary sample on the lines suggested by Sukhatme was impracticable, we should use proportionate sampling, and the size of S would usually ensure that $\text{var}(m_p) < \text{var}(\bar{x})$. In the second case, we should use optimum stratified sampling, and rely on the variability of the σ_k to ensure that $\text{var}(m_o) < \text{var}(\bar{x})$. Since an adequate degree of knowledge about the σ_k would be unlikely unless the N_k were quite large, we should in this case almost certainly be safe in using the method. To the above considerations must be added the fact that if very inaccurate estimates of the σ_k are used in (iii), then, whatever the nature of the population, the resulting procedure may be extremely inefficient.

It must be realized, of course, that even if it were known that $\text{var}(m_o) < \text{var}(\bar{x})$, it would not follow that the optimum stratified method would necessarily be the most convenient. It may be impossible, or at any rate inconvenient, to do any sort of random sampling, and some sort of quasi-random sampling may have to be used (see. e.g. Madow & Madow, 1944), but if the principle of random sampling is applicable the stratified method is not likely to be much more inconvenient, and in fact in most cases will be more convenient, than the unrestricted method.

Summary

The stratified method has been used in the past almost solely for large-scale social and agricultural surveys. Here the stratum sizes are large, and known results for infinite populations apply. There seems no reason why stratified sampling should not be used to advantage for smaller populations, and it is important to know to what extent these results still apply. In this paper a comparison has been made with unrestricted random sampling in the usual case where we are interested in estimating the mean. The advantages of the stratified method are modified, but in most cases where the method is applicable it will be found to be worth while.

The above work was carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory. The author desires to acknowledge the assistance rendered by Mr D. V. Lindley who prepared the diagrams.

References

Bowley, A. L. (1926). Measurement of the precision attained in sampling. *Bull. Int. Inst. Statist.* **22**, 1ère livraison.

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, **1**. Griffin and Co.

Kendall, M. G. (1946). *The Advanced Theory of Statistics*, **2**. Griffin and Co.

Madow, W. G. & Madow, L. H. (1944). On the theory of systematic sampling. *Ann. Math. Statist.* **15**, 1–24.

Neyman, J. (1934). On the two different aspects of the representative method. *J. Roy. Statist. Soc.* **91**, 558–625.

Sukhatme, P. V. (1935). Contribution to the theory of the representative method. *Suppl. J. Roy. Statist. Soc.* **2**, 253–68.

Wilks, S. S. (1943). *Mathematical Statistics*. Princeton University Press.

SOME THEOREMS ON TIME SERIES. I

By P. A. P. MORAN
Institute of Statistics, Oxford University

One of the principal problems in the theory of time series is to discuss the relation between two series, and in the present paper we prove a theorem by which we can test whether two such series are independent. Such a test of significance must depend on the models which we assume for the probability processes which generate the series. In practice, the two most useful models are, first, that of a moving average of a series of independent random components and, secondly, the solutions of linear stochastic difference equations.

Let

$$\ldots, \eta(t-1), \eta(t), \eta(t+1), \ldots$$

be a sequence of independent random variables each distributed in the same distribution which we take to have zero mean and its second, third, and fourth moments finite. Then the time series generated by

$$X(t) = \sum_{i=0}^{N} \alpha_i \eta(t-i)$$

is a moving average with weights α_i. On the other hand, consider a stochastic difference equation of the form

$$X(t) + a_1 X(t-1) + \ldots + a_h X(t-h) = \eta(t). \tag{1}$$

In order that the solution of (1) for successive values of t shall form a stationary series it is necessary to impose the condition that the roots of the characteristic equation

$$z^h + a_1 z^{h-1} + \ldots + a_h = 0 \tag{2}$$

shall all lie inside the circle $|z| = 1$ (Wold, 1938, p. 53). When this is true the solution of (1) can be shown to be of the form

$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i),$$

where the α_i are certain functions of the roots of (2). In this case $\sum_{i=0}^{\infty} |\alpha_i|$ is majorized by a convergent geometric series.

Thus we see that both the above models are included in the more general one in which we define $X(t)$ as given by

$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i),$$

where the α_i are any sequence of constants satisfying $\sum_{i=0}^{\infty} |\alpha_i| < \infty$. Now suppose

$$\ldots, \zeta(t-1), \zeta(t), \zeta(t+1), \ldots$$

is another sequence of independent random variables having a distribution with zero mean and finite second, third and fourth moments. We write

$$Y(t) = \sum_{i=0}^{\infty} \beta_i \zeta(t-i),$$

where $\sum_{i=0}^{\infty} |\beta_i| < \infty$. To discuss whether two such empirical series of this form are correlated we prove that the covariance

$$S = \sum_{1}^{n} X(t)\, Y(t) \tag{3}$$

tends, as n increases, to be distributed in the normal form about zero mean with a second moment which is a function of the α_i and the β_i. We shall discuss later the calculation of this second moment from empirical series, in which case some care is necessary.

We first illustrate our method of proof by considering the much simpler problem of determining the asymptotic distribution of the sum

$$T_n = \sum_{s=1}^{n} X(t-s). \tag{4}$$

We shall show that this asymptotic distribution is also, under certain conditions, normal. This result is interesting because it establishes a central limit theorem (and therefore a law of large numbers) for stationary stochastic processes of this type. The law of large numbers for Markov chains has been considered by several writers, in particular Bernstein (1927), who proves his results by using central limit theorems for non-independent components. His theorems cannot be applied in the present case, but some of the ideas of his methods can.

Consider (4) above, where $X(t)$ is defined by

$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i)$$

and $\sum_{i=0}^{\infty} |\alpha_i|$ is convergent. There is no loss in generality in supposing that

$$\sum_{i=0}^{\infty} |\alpha_i| < 1.$$

Clearly
$$E(T_n) = \sum_{s=1}^{n} \sum_{i=0}^{\infty} \alpha_i E[\eta(t-s-i)]$$
$$= 0.$$

Write $\sigma^2 = E(\eta^2), \quad c_0 = E[X(t)^2], \quad c_s = E[X(t)\,X(t-s)].$

Then $c_0 = \sigma^2(\alpha_0^2 + \alpha_1^2 + \ldots), \quad c_s = \sigma^2(\alpha_0\alpha_s + \alpha_1\alpha_{s+1} + \ldots),$

which are both clearly convergent. Moreover,

$$R_n = E(T_n^2) = nE[X(t)^2] + 2\sum_{i=1}^{n-1} \sum_{s=1}^{n-i} E[X(t-i)\,X(t-i-s)]$$
$$= \left(nc_0 + 2\sum_{i=1}^{n-1} (n-i)\,c_i\right).$$

$n^{-1}R_n$ tends, as n increases, to $\quad R_0 = \left(c_0 + 2\sum_{i=1}^{\infty} c_i\right)$

if this series converges absolutely. We shall show that $\lim n^{-1}R_n$ is finite. For R_0 is clearly

$$\left(\sum_{i=0}^{\infty} \alpha_i\right)^2 \sigma^2, \tag{5}$$

and this is finite. Moreover, we notice that $n^{-1}R_n$ is not greater than

$$\left(\sum_{i=0}^{\infty} |\alpha_i|\right)^2 \sigma^2.$$

We must now impose the condition that $\sum_{i=0}^{\infty} \alpha_i$ is not zero. This condition is necessary to our

method of argument. If it is not zero, it may be assumed, without loss of generality, greater than a positive number. We now show that as n increases

$$pr\{t_0(2R_n)^{\frac{1}{2}} \leqslant T_n < t_1(2R_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\,dt.$

We require the following lemma (Bernstein, 1927, p. 12):

LEMMA I. Let $$\rho_n = \Sigma_n + \sigma_n,$$

where ρ_n, Σ_n and σ_n are random variables such that

$$E(\Sigma_n) = E(\sigma_n) = 0, \quad E(\Sigma_n^2) = H_n, \quad E(\sigma_n^2) = H_n'.$$

Then if, for n large, $$pr\{t_0(2H_n)^{\frac{1}{2}} \leqslant \Sigma_n < t_1(2H_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\,dt,$

then $$pr\{t_0(2J_n)^{\frac{1}{2}} \leqslant \rho_n < t_1(2J_n)^{\frac{1}{2}}\},$$

where $J_n = E(\rho_n^2)$ tends, uniformly in t_0 and t_1, to

$$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\,dt,$$

provided that $$\lim_{n\to\infty} \frac{H_n'}{H_n} = 0.$$

Let ϵ be an arbitrarily small number and choose N so large that

$$\sum_{i=N}^{\infty} |\alpha_i| < \epsilon \sum_{i=0}^{\infty} |\alpha_i| < \epsilon.$$

Write $$X_1(t) = \sum_{i=0}^{N} \alpha_i \eta(t-i), \quad T_n' = \sum_{s=1}^{n} X_1(t-s).$$

Then $$E(T_n') = 0,$$

and write $$R_n' = E(T_n'^2).$$

We shall prove that the distribution of T_n' tends to normality, i.e. that

$$pr\{t_0(2R_n')^{\frac{1}{2}} \leqslant T_n' < t_1(2R_n')^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\,dt.$

We first calculate R_n and R_n' in another way. For

$$\begin{aligned} T_n &= \sum_{s=1}^{n} X(t-s) = \sum_{s=1}^{n} \sum_{i=0}^{\infty} \alpha_i \eta(t-s-i) \\ &= \alpha_0 \eta(t-1) + (\alpha_0+\alpha_1)\eta(t-2) + \ldots + (\alpha_0+\ldots+\alpha_{n-1})\eta(t-n) \\ &\quad + \sum_{s=1}^{\infty} (\alpha_s + \ldots + \alpha_{s+n-1})\eta(t-s-n), \end{aligned}$$

and so $$R_n = E(T_n^2) = \sigma^2\left\{\alpha_0^2 + (\alpha_0+\alpha_1)^2 + \ldots + (\alpha_0+\ldots+\alpha_{n-1})^2 + \sum_{s=1}^{\infty} (\alpha_s+\ldots+\alpha_{s+n-1})^2\right\},$$

and this series converges. On the other hand,

$$T'_n = \alpha_0\eta(t-1)+(\alpha_0+\alpha_1)\eta(t-2)+\ldots+(\alpha_0+\ldots+\alpha_{N-1})\eta(t-N)$$
$$+\sum_{p=1}^{n-N}(\alpha_0+\ldots+\alpha_N)\eta(t-N-p)$$
$$+(\alpha_1+\ldots+\alpha_N)\eta(t-n-1)+\ldots+\alpha_N\eta(t-n-N),$$

and so
$$R'^2_n = \sigma^2\{\alpha_0^2+\ldots(\alpha_0+\alpha_1)^2+\ldots+(\alpha_0+\ldots+\alpha_{N-1})^2$$
$$+(n-N)(\alpha_0+\ldots+\alpha_N)^2+(\alpha_1+\ldots+\alpha_N)^2+\ldots+\alpha_N^2\}.$$

Since we have already supposed that $\sum_{i=0}^{\infty}\alpha_i$ is positive, there exist positive numbers N_0 and d such that for all $N>N_0$, $\sum_{i=0}^{N}\alpha_i>d$. If this is not true the theorem is in general false. For suppose the distribution of the η's to be non-normal and write $\alpha_0=1$, $\alpha_1=-1$, $\alpha_i=0$ $(i>1)$. Then the distribution of T_n does not tend to normality and its variance does not increase with n. We shall later show that this condition on the α's is in fact satisfied for the solutions of stochastic difference equations.

Now by the ordinary central limit theorem, as n increases,

$$T''_n = \sum_{p=1}^{n-N}(\alpha_0+\ldots+\alpha_N)\eta(t-N-p)$$

tends to be distributed normally with zero mean and variance

$$R''_n = (n-N)(\alpha_0+\ldots+\alpha_N)^2\sigma^2,$$

that is
$$pr\{t_0(2R''_n)^{\frac{1}{2}} \leqslant T''_n < t_1(2R''_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1}e^{-t^2}\,dt.$$

Using Lemma I we see that the same is true, for fixed N, when we replace T''_n by T'_n and R''_n by R'_n. Now

$$T_n = T'_n+Q,$$

say, where Q is what we get if we replace the sequence $(\alpha_0,\alpha_1,\ldots)$ in T_n by $(\alpha_{N+1},\ldots)$ and alter t, and from (5) we can choose N so large that for $n>N$, $n^{-1}E(Q^2)<\epsilon$, say. Taking a sequence ϵ_1, ϵ_2, ... tending to zero and choosing first N sufficiently large and then n and using Lemma I again, we see that

$$pr\{t_0(2R_n)^{\frac{1}{2}} \leqslant T_n < t_1(2R_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1}e^{-t^2}\,dt.$$

To complete the discussion we must show that the condition we have imposed on the sequence $\alpha_0,\alpha_1,\ldots$ is satisfied by the coefficients of the solutions of stationary stochastic difference equations. Consider an equation

$$X(t)+a_1X(t-1)+\ldots+a_hX(t-h)=\eta(t)$$

such that the roots of
$$z^h+a_1z^{h-1}+\ldots+a_h=0 \qquad (6)$$

all lie inside the circle $|z|=1$. Then the solution of this equation is given (Wold, 1938, p. 53) by

$$X(t)=\sum_{i=0}^{\infty}\alpha_i\eta(t-i),$$

where the α_i are now the solutions of the infinite set of equations

$$\begin{aligned}
\alpha_0 &= 1 \\
a_1\alpha_0+\alpha_1 &= 0, \\
a_2\alpha_0+a_1\alpha_1+\alpha_2 &= 0, \\
&\cdots\cdots \\
a_h\alpha_0+a_{h-1}\alpha_1+\ldots+\alpha_h &= 0, \\
a_h\alpha_1+\ldots+a_1\alpha_h+\alpha_{h+1} &= 0, \\
&\cdots\cdots
\end{aligned}$$

and since the left-hand side is an absolutely convergent double series, we add, obtaining

$$(1+a_1+\ldots+a_h)\sum_{i=0}^{\infty}\alpha_i = 1,$$

and so $\sum_{i=0}^{\infty}\alpha_i \neq 0$ and, as already observed, without loss of generality, may be supposed positive. This quantity is finite because all the roots of equation (6) lie inside the circle $|z| = 1$. Moreover, it follows that

$$R_0 = \left(\sum_{i=0}^{\infty}\alpha_i\right)^2\sigma^2 = (1+a_1+\ldots+a_n)^{-2}\sigma^2.$$

This is, in fact, proportional to the derivative at zero of the integrated power spectrum (Wold, 1938, p. 69).

We now turn to the problem of discussing the relation between two such series and we consider the asymptotic distribution of

$$S_n = \sum_{t=1}^{n} X(-t)\,Y(-t),$$

where
$$X(t) = \sum_{i=1}^{\infty}\alpha_i\eta(t-i) \quad (\alpha_1 \neq 0), \tag{7}$$

and
$$Y(t) = \sum_{i=1}^{\infty}\beta_i\zeta(t-i) \quad (\beta_1 \neq 0). \tag{8}$$

We write S_n in this form rather than that of (3) for the sake of convenience in what follows, and we have altered the notation of the sums (7) and (8) so that they begin with the coefficients α_1 and β_1 for the same reason. Writing

$$c_s = E[X(t)\,X(t-s)], \quad d_s = E[Y(t)\,Y(t-s)] \quad (s = 0, 1, \ldots),$$

as before, we have

$$c_s = \sigma_1^2(\alpha_1\alpha_{s+1}+\alpha_2\alpha_{s+2}+\ldots), \quad d_s = \sigma_2^2(\beta_1\beta_{s+1}+\beta_2\beta_{s+2}+\ldots),$$

where σ_1^2 and σ_2^2 are the second moments of η and ζ. Then

$$E(S_n) = \sum_{t=1}^{n} E(X(-t)\,Y(-t)) = 0,$$

$$\begin{aligned}
E(S_n^2) &= E\left\{\sum_{t=1}^{n} X(-t)\,Y(-t)\right\}^2 \\
&= E\left\{\sum_{t=1}^{n} X^2(-t)\,Y^2(-t) + 2\sum_{t=1}^{n-1}\sum_{s=1}^{n-t} X(-t)\,X(-t-s)\,Y(-t)\,Y(-t-s)\right\} \\
&= nc_0d_0 + 2\sum_{s=1}^{n-1}(n-s)\,c_sd_s.
\end{aligned} \tag{9}$$

Consider the behaviour of $n^{-1}E(S_n^2)$ as n increases. Clearly

$$n^{-1}E(S_n^2) \to c_0 d_0 + 2\sum_{s=1}^{\infty} c_s d_s = C, \quad \text{say,} \tag{10}$$

if the series C is absolutely convergent. If X and Y are moving averages or the solutions of stationary stochastic difference equations this is certainly true, for in the first case the series is finite, and in the second it is majorized by a convergent geometric series. We show that it is true in the general case by the following argument. Without restricting generality, we may assume, as before, that $\sum_1^{\infty} |\alpha_i| < 1, \sum_1^{\infty} |\beta_i| < 1$. Then

$$\begin{aligned} |c_s| &\leqslant \sigma_1^2(|\alpha_1\alpha_s| + \ldots) \\ &\leqslant \sigma_1^2(|\alpha_1| + \ldots), \end{aligned}$$

and so

$$\begin{aligned} \left|\sum_1^{\infty} c_s d_s\right| &\leqslant \sum_1^{\infty} |c_s d_s| \leqslant \sigma_1^2 \sum_1^{\infty} |d_s| \\ &\leqslant \sigma_1^2\sigma_2^2 \sum_{s=1}^{\infty} (|\beta_1||\beta_s| + \ldots) \\ &\leqslant \sigma_1^2\sigma_2^2 \left(\sum_{i=1}^{\infty} |\beta_i|\right)^2. \end{aligned} \tag{11}$$

Also

$$c_0 d_0 \leqslant \sigma_1^2\sigma_2^2 \left(\sum_{i=1}^{\infty} |\alpha_i|^2\right)\left(\sum_{i=1}^{\infty} |\beta_i|^2\right),$$

and so $c_0 d_0 + 2\sum_1^{\infty} c_s d_s$ is finite. We now prove that C is not zero. For

$$C = c_0 d_0 + 2\sum_{s=1}^{\infty} c_s d_s = \sigma_1^2\sigma_2^2\left[(\alpha_1^2 + \alpha_2^2 + \ldots)(\beta_1^2 + \beta_2^2 + \ldots) + 2\sum_{s=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \alpha_m \alpha_{m+s} \beta_n \beta_{n+s}\right],$$

and after some rearrangement, this equals

$$\sigma_1^2\sigma_2^2[(\alpha_1\beta_1)^2 + (\alpha_1\beta_2 + \alpha_2\beta_1)^2 + (\alpha_1\beta_3 + \alpha_2\beta_2 + \alpha_3\beta_1)^2 + \ldots],$$

and $(\alpha_1\beta_1)^2$ is greater than zero and the rest non-negative at least. We therefore conclude that

$$n^{-1}E(S_n^2) \to C,$$

where

$$0 < C < \infty.$$

Assuming as before that

$$\sum_1^{\infty} |\alpha_i| < 1, \quad \sum_1^{\infty} |\beta_i| < 1,$$

we define N so that

$$\sum_{N+1}^{\infty} |\alpha_i| < \epsilon \sum_1^{\infty} |\alpha_i| < \epsilon, \tag{12}$$

$$\sum_{N+1}^{\infty} |\beta_i| < \epsilon \sum_1^{\infty} |\beta_i| < \epsilon, \quad \text{where } \epsilon \text{ is small.} \tag{13}$$

We now write

$$X_1(t) = \sum_{i=1}^{N} \alpha_i \eta(t-i), \tag{14}$$

$$Y_1(t) = \sum_{i=1}^{N} \beta_i \zeta(t-i), \tag{15}$$

and consider the sum

$$S_n' = \sum_1^n X_1(-t) Y_1(-t). \tag{16}$$

We begin by proving that when n is large this sum tends to be distributed in the normal form with a variance which is asymptotically equal to nC_1, where C_1 is obtained from C by putting $\alpha_i = \beta_i = 0$ for $i > N$. For it is then clearly true that

$$n^{-1}E(S_n'^2) \to C_1.$$

Now consider

$$S_n' = \sum_1^n X_1(-t)\,Y_1(-t),$$

where n is greater than N. For convenience of notation, we write

$$\eta_i = \eta(-i), \quad \zeta_i = \zeta(-i).$$

We then have

$$S_n' = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}\eta_i\zeta_j,$$

where the A_{ij} are certain constants. Moreover

$$\begin{aligned}
E(\eta_i\zeta_j) &= 0 && \text{all } i, j,\\
E(\eta_i^2\zeta_j^2) &= \sigma_1^2\sigma_2^2 && \text{all } i, j,\\
E(\eta_i^2\zeta_j\zeta_k) &= E(\eta_j\eta_k\zeta_l^2) = 0 && \text{for } j \neq k,\\
E(\eta_i\eta_j\zeta_k\zeta_l) &= 0 && \text{if } i \neq j \text{ or } k \neq l.
\end{aligned}$$

It therefore follows that

$$E(S_n'^2) = \sigma_1^2\sigma_2^2\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}^2.$$

Inserting (14) and (15) in (16) we have

$$A_{ij} = 0$$

if

$$i > n+N, \quad \text{or} \quad j > n+N \quad \text{or} \quad |i-j| > N-1,$$

and

$$S_n' = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4,$$

where

$$\Sigma_1 = \sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}\eta_i\zeta_j$$

with

$$\begin{aligned}
A_{ij} &= \alpha_i\beta_j + \alpha_{i-1}\beta_{j-1} + \ldots + \alpha_{1+i-j}\beta_1 && \text{for } i > j,\\
&= \alpha_i\beta_j + && + \alpha_1\beta_{1+j-i} && \text{for } i < j,\\
&= \alpha_1\beta_1 + && + \alpha_i\beta_j && \text{for } i = j.
\end{aligned}$$

We also have

$$\Sigma_2 = \Sigma\Sigma A_{ij}\eta_i\zeta_j,$$

where the sum is taken over values of i and j such that $|i-j| < N$, $i \leqslant n$, $j \leqslant n$ and either $N < i$ or $N < j$, where

$$\begin{aligned}
A_{ij} &= \alpha_1\beta_{p+1} + \ldots + \alpha_{N-p}\beta_N && \text{for } j - i = p > 0\\
&= \alpha_{p+1}\beta_1 + \ldots + \alpha_N\beta_{N-p} && \text{for } i - j = p > 0\\
&= \alpha_1\beta_1 + \ldots + \alpha_N\beta_N && \text{for } i = j.
\end{aligned}$$

Then

$$E(\Sigma_2) = 0, \quad E(\Sigma_2^2) = \sigma_1^2\sigma_2^2\Sigma\Sigma A_{ij}^2,$$

where the sum is taken over the above values of i and j. This equals

$$\begin{aligned}
(n-N)\sigma_1^2\sigma_2^2[&(\alpha_1\beta_N)^2 + (\alpha_1\beta_{N-1} + \alpha_2\beta_N)^2 + \ldots\\
&+ (\alpha_1\beta_1 + \ldots + \alpha_N\beta_N)^2 + \ldots + (\alpha_N\beta_2 + \alpha_{N-1}\beta_1)^2 + (\alpha_N\beta_1)^2]. \qquad (17)
\end{aligned}$$

We know that $\alpha_1 \neq 0$. Let β_t be the first term of the sequence $\beta_N, \beta_{N-1}, \ldots, \beta_1$ which is not zero. Such a term certainly exists. Then the sum in the outer brackets of (17) will contain a term of the form $(\alpha_1\beta_t)^2$ and consequently $E(\Sigma_2^2) > 0$, and for N fixed will increase as $(n-N)$.

Next we have

$$\Sigma_3 = \Sigma\Sigma A_{ij}\eta_i\zeta_j,$$

where either $$i \leqslant n, j > n \quad \text{and} \quad j - i < N, \tag{18}$$

or $$j \leqslant n, i > n \quad \text{and} \quad i - j < N, \tag{19}$$

and
$$\begin{aligned} A_{ij} &= \alpha_{p+1}\beta_1 + \ldots + \alpha_N\beta_{N-p} \quad \text{for} \quad i-j = p > 0 \\ &= \alpha_1\beta_{p+1} + \ldots + \alpha_{N-p}\beta_N \quad \text{for} \quad j-i = p > 0. \end{aligned}$$

Then $$E(\Sigma_3) = 0, \quad E(\Sigma_3^2) = \sigma_1^2\sigma_2^2\Sigma\Sigma A_{ij}^2,$$

where the sum is taken over the values (18) and (19).

Finally $$\Sigma_4 = \sum_{i=n+1}^{n+N}\sum_{j=n+1}^{n+N} A_{ij}\eta_i\zeta_j,$$

where
$$\begin{aligned} A_{ij} &= \alpha_{i-n}\beta_{i-p-n} + \ldots + \alpha_N\beta_{N-p} \quad \text{for} \quad i-j = p > 0 \\ &= \alpha_{j-p-n}\beta_{j-n} + \ldots + \alpha_{N-p}\beta_N \quad \text{for} \quad j-i = p > 0 \\ &= \alpha_p\beta_p \qquad + \ldots + \alpha_N\beta_N \quad \text{for} \quad i = j = n+p > n, \end{aligned}$$

and $$E(\Sigma_4) = 0, \quad E(\Sigma_4^2) = \sigma_1^2\sigma_2^2 \sum_{i=n+1}^{n+N}\sum_{j=n+1}^{n+N} A_{ij}^2.$$

We readily see that $$E(\Sigma_i\Sigma_j) = 0 \quad \text{for} \quad i \neq j$$

and therefore $$E(S_n'^2) = E(\Sigma_1^2) + E(\Sigma_2^2) + E(\Sigma_3^2) + E(\Sigma_4^2).$$

Moreover, for constant N, $E(\Sigma_1^2)$, $E(\Sigma_2^2)$ and $E(\Sigma_3^2)$ are constant, and so for large n we have

$$\begin{aligned} n^{-1}E(S_n'^2) \to C_2 = \sigma_1^2\sigma_2^2[&(\alpha_1\beta_N)^2 + (\alpha_1\beta_{N-1} + \alpha_2\beta_N)^2 + \ldots \\ &+ (\alpha_1\beta_1 + \ldots + \alpha_N\beta_N)^2 + \ldots + (\alpha_N\beta_1)^2] \neq 0. \end{aligned} \tag{20}$$

Now suppose that N is fixed and consider the sum $\sum_1^n X_1(-t)\,Y_1(-t)$. We write

$$n = m(m+N) + p,$$

where $p < 2m+N+1$ and n is large enough for m to be greater than N. This equation fixes m which increases roughly as $n^{\frac{1}{2}}$ when n increases. Write

$$\begin{aligned} S_n' &= \left(\sum_{t=1}^{N} + \sum_{N+1}^{N+m} + \ldots + \sum_{1+m(N+m-1)}^{m^2+mN} + \sum_{n-p+1}^{n}\right) X_1(-t)\,Y_1(-t) \\ &= V_1 + U_1 + V_2 + U_2 + \ldots + V_m + U_m + W. \end{aligned}$$

Then $V_1, \ldots, V_m$ and W are all independent and $E(V_1^2), \ldots, E(V_m^2)$ are independent of n, and in fact not greater than KN, where K is a constant independent of N. Also $E(W^2)$ is not greater than $K(2m+N+1)$, where K may be taken as the same constant. $U_1, \ldots, U_m$ are also all independent and $E(U_m^2)$ is asymptotically equal to mC_2 when n (and therefore m) are large. Therefore, writing

$$A_m = U_1 + \ldots + U_m, \quad B_m = V_1 + \ldots + V_m + W,$$

we have $$E(A_m) = 0, \quad E(A_m^2) = \sum_{i=1}^{m} E(U_i^2),$$

$$E(B_m) = 0, \quad E(B_m^2) = \sum_{i=1}^{m} E(V_i^2) + E(W^2),$$

and the latter increases as m, whilst the former increases as m^2 and so

$$E(B_m^2)\{E(A_m^2)\}^{-1} \to 0$$

as n increases.

By Lemma I it is therefore sufficient to show that the distribution of A_m tends to normality.

LEMMA II (Liapounoff's Central Limit Theorem, Bernstein, 1927). If

$$\Sigma_m = u_1^{(m)} + \ldots + u_m^{(m)}$$

is the sum of m independent quantities such that

$$E(u_r^{(m)}) = 0, \quad E(u_r^{(m)2}) = b_r^{(m)}, \quad E(u_r^{(m)4}) = c_r^{(m)},$$

and if, as m increases, $$b_m^{-2} \sum_{r=1}^{m} c_r^{(m)} \to 0,$$

where $$b_m = \sum_{r=1}^{m} b_r^{(m)} = E(\Sigma_m^2),$$

then $$pr\{t_0(2b_m)^{\frac{1}{2}} \leqslant \Sigma_m < t_1(2b_m)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt.$$

To apply the lemma we put $U_r = u_r^{(m)}$. We already have $E(U_r) = 0$. Also

$$m^{-1}E(U_r^2) \to C_2 > 0 \quad \text{by (20)},$$

and so $$m^{-2}b_m \to C_2.$$

Now consider $$c_r^{(m)} = E(U_r^4),$$

where $$U_r = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} A_{pq}\, \eta_{r(N+r-1)+p-1}\, \zeta_{r(N+r-1)+q-1},$$

and the A_{pq} are calculated with m in place of n. Since the η's all have the same probability distribution and similarly for the ζ's, we shall write η_p and ζ_q for $\eta_{r(N+r-1)+p-1}$ and $\zeta_{r(N+r-1)+q-1}$ for the sake of convenience. So we can write the above

$$U_r = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} A_{pq}\, \eta_p\, \zeta_q.$$

U_r^4 will be a polynomial of the fourth order in the η's and the ζ's and its expectation may be regarded as the sum of two distinct types of terms so that $E(U_r^4) = \Sigma E(w_1) + \Sigma E(w_2)$, where the terms w_1 are of the form $A_{pq}^4 \eta_p^4 \zeta_q^4$, and the terms w_2 are of the form $A_{pq}^2 A_{kl}^2 \eta_p^2 \zeta_q^2 \eta_k^2 \zeta_l^2$ with $(p, q) \neq (k, l)$. All other terms arising in the product will clearly vanish when the expectation is taken.

Then, since the A_{pq} are bounded and the number of non-zero terms in w_1 and w_2 are not greater than $2N(m+N)$ and $4N^2(m+N)^2$ respectively, we have

$$E(U_r^4) < Km^2,$$

where K is a constant depending on N but independent of m and n. It follows that

$$b_m^{-2} \sum_{1}^{m} C_r^{(m)}$$

is of order m^{-1} and tends to zero as n and m increase. The conditions of the lemma are therefore satisfied and we conclude that

$$pr\{t_0(2E(A_m^2))^{\frac{1}{2}} \leqslant A_m < t_1(2E(A_m^2))^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $\quad \pi^{-\frac{1}{2}}\int_{t_0}^{t_1} e^{-t^2}\,dt.$

Applying Lemma I we have

$$pr\{t_0[2E(S_n'^2)]^{\frac{1}{2}} \leqslant S_n' < t_1[2E(S_n'^2)]^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to the same limit.

We now consider the relationship between S_n' and S_n. Write

$$\begin{aligned}
S_n'' = S_n - S_n' &= \sum_1^n X(-t)\,Y(-t) - \sum_1^n X_1(-t)\,Y_1(-t) \\
&= \sum_{t=1}^{n}\left(\sum_{i=1}^{\infty}\alpha_i\eta_{t+i}\right)\left(\sum_{j=1}^{\infty}\beta_j\xi_{t+j}\right) - \sum_{t=1}^{n}\left(\sum_{i=1}^{N}\alpha_i\eta_{t+i}\right)\left(\sum_{j=1}^{N}\beta_i\zeta_{t+j}\right) \\
&= \sum_{t=1}^{n}\left(\sum_{i=-N+1}^{\infty}\alpha_i\eta_{t+i}\right)\left(\sum_{j=-N+1}^{\infty}\beta_j\zeta_{t+j}\right) \\
&\quad + \sum_{t=1}^{n}\left(\sum_{i=-N+1}^{\infty}\alpha_i\eta_{t+i}\right)\left(\sum_{j=1}^{N}\beta_j\zeta_{t+j}\right) \\
&\quad + \sum_{t=1}^{n}\left(\sum_{i=1}^{N}\alpha_i\eta_{t+i}\right)\left(\sum_{j=-N+1}^{\infty}\beta_j\zeta_{t+j}\right) \\
&= W_1 + W_2 + W_3. \qquad (21)
\end{aligned}$$

We must now calculate the variance of these terms. Consider again (9). We have shown (11) that

$$\begin{aligned}
2\sum_1^{\infty} c_s d_s &\leqslant 2\sigma_1^2\sigma_2^2(|\beta_0| + |\beta_1| + \ldots)^2 \\
&\leqslant 2\sigma_1^2\sigma_2^2(|\alpha_0| + |\alpha_1| + \ldots)^2,
\end{aligned}$$

and we now apply this to the three sums in equation (21). It follows that if N be chosen to satisfy the conditions (12) and (13) then

$$\overline{\lim}\, n^{-1}E(W_1^2) < K\sigma_1^2\sigma_2^2\epsilon^2, \quad \overline{\lim}\, n^{-1}E(W_2^2) < K\sigma_1^2\sigma_2^2\epsilon^2, \quad \overline{\lim}\, n^{-1}E(W_3^2) < K\sigma_1^2\sigma_2^2\epsilon^2,$$

where K is a constant independent of N.

It follows that

$$S_n = S_n' + W_1 + W_2 + W_3,$$

where the variance of W_1, W_2 and W_3 can be made small compared with that of S_n by choosing N large. Then by first choosing N large and then n and using Tchebycheff's inequality, we see that the distribution of S_n tends to normality with variance $E(S_n^2)$ and this completes the proof.

In the general application of the above results some care is needed. We can suppose that our empirical values of X and Y are distributed about their sample means which we take to be zero and we must estimate the variance of S_n from formulae (9), or (approximately) from (10). But we must not insert in this formula the sample covariances for the c_s and the d_s because, as Bartlett (1946) has shown, the standard errors of the sample values of these covariances are of order $n^{-\frac{1}{2}}$ and we cannot therefore expect the series (10) to converge, let

alone give the correct value. To use the formula correctly we must first decide on the order and coefficients of the stochastic difference equation which we can suppose generated the series and, from these coefficients, calculate the value of (9).

In the case where the series are generated by a three-term difference equation, the calculations are simplified. Suppose the X and Y satisfy the equations

$$X(t+2)+aX(t+1)+bX(t)=\eta(t+2),$$

$$Y(t+2)+AY(t+1)+BY(t)=\zeta(t+2),$$

where $$E(\eta(t))=E(\zeta(t))=0$$

and $$E(\eta^2(t))=\sigma_1^2, \quad E(\zeta^2(t))=\sigma_2^2,$$

as before. For the series to be stationary, we must have $b<1$, $B<1$. We suppose that in addition to this the series are oscillatory and so $a^2<4b$, $A^2<4B$. The solutions will then be

$$X(t)=\sum_{s=0}^{\infty} 2(4b-a^2)^{-\frac{1}{2}}p^s\sin\theta s\,\eta(t-s+1),$$

$$Y(t)=\sum_{s=0}^{\infty} 2(4B-A^2)^{-\frac{1}{2}}P^s\sin\phi s\,\zeta(t-s+1),$$

where $p=b^{\frac{1}{2}}$, $P=B^{\frac{1}{2}}$, $\cos\theta=-a(2b^{\frac{1}{2}})^{-1}$, $\cos\phi=-A(2B^{\frac{1}{2}})^{-1}$. Also (Kendall, 1946, p. 408)

$$r_s=\frac{c_s}{c_0}=\frac{p^s\sin(s\theta+\psi)}{\sin\psi}, \quad R_s=\frac{d_s}{d_0}=\frac{P^s\sin(s\phi+\Phi)}{\sin\Phi},$$

where $$\tan\psi=\frac{1-p^2}{1+p^2}\tan\theta \quad \text{and} \quad \tan\Phi=\frac{1-P^2}{1+P^2}\tan\phi$$

and $$c_0=\sigma_1^2\frac{1+b}{(1-b)\{(1+b)^2-a^2\}}, \quad d_0=\sigma_2^2\frac{1+B}{(1-B)\{(1+B)^2-A^2\}}.$$

We then need to calculate

$$\begin{aligned}
C&=c_0d_0+2\sum_{s=1}^{\infty}c_sd_s\\
&=c_0d_0\left\{1+2\sum_{s=1}^{\infty}r_sR_s\right\}\\
&=c_0d_0\left\{1+\sum_{s=1}^{\infty}\frac{p^sP^s\sin(s\theta+\psi)\sin(s\phi+\Phi)}{\sin\psi\sin\Phi}\right\}\\
&=c_0d_0\left\{1+\frac{2pP}{\sin\psi\sin\Phi}\left[\frac{\cos(\psi-\Phi+\theta-\phi)-pP\cos(\psi-\Phi)}{1-2pP\cos(\theta-\phi)+p^2P^2}\right.\right.\\
&\qquad\left.\left.-\frac{\cos(\psi+\Phi+\theta+\phi)-pP\cos(\psi+\Phi)}{1-2pP\cos(\theta+\phi)+p^2P^2}\right]\right\}. \qquad (22)
\end{aligned}$$

It is probably easiest to calculate C from this equation rather than attempt to simplify (22) still further. I hope to discuss the practical application of these formulae in another paper.

REFERENCES

WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala.

BERNSTEIN, S. (1927). *Math. Ann.* **97**, 1.

BARTLETT, M. S. (1946). *Supp. J. Roy. Statist. Soc.* **8**, 27.

KENDALL, M. G. (1946). *Advanced Theory of Statistics*, 2. London: Charles Griffin and Co.

RANK CORRELATION BETWEEN TWO VARIABLES, ONE OF WHICH IS RANKED, THE OTHER DICHOTOMOUS

By J. W. WHITFIELD, *Psychological Laboratory, University of Cambridge*

Rank correlation is one of the most useful statistical techniques available for the treatment of data arising in experimental and applied psychological research. Chambers (1946) has indicated the type of data most frequently occurring in these fields, and has pointed out the advantages of Kendall's τ over Spearman's ρ or any form of transformation to ordinal form.

Given the use of τ when tied rankings are present (Kendall, 1946) it seemed possible to extend the method to cover a very common problem in psychology, namely, determination of the relation between two variables, one of which is expressed as a ranking and the other as a dichotomy. In applied or field work the relation of a psychological 'measurement' and an external criterion nearly always appears in this form. The usual method of determining the relationship consists of reducing the ranking to a dichotomy and calculating χ^2 for the 2×2 table which results. That this may lead to inaccuracy can be seen from the following example:

Variable A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Variable B	+	+	+	+	+	+	+	−	−	−	+	+	+	−	−	−	−	−	−	−
Variable C	−	−	−	+	+	+	+	+	+	+	−	−	−	−	−	−	−	+	+	+

Here the data are supposed to be ranked according to variable A and dichotomized into + and − with respect to variables B and C.

Treating the relation between variables A and B as a 2×2 contingency table:

	Variable B	
	+	−
Variable A: Rankings 1–10	7	3
Rankings 11–20	3	7

Applying χ^2, P is found to be 0·074 without Yates's correction for continuity, or 0·180 if the correction is applied.

But χ^2 is exactly the same for the contingency table relating variables A and C, although it is obvious from the data that there is considerable difference in the two relationships, the evidence for which is sacrificed by reducing the ranking to a dichotomy.

If, alternatively, we consider the dichotomous variable as a ranking composed entirely of two sets of tied rankings, we may calculate the coefficients between A and B, A and C respectively which I shall denote by τ_{AB}, τ_{AC}. The corresponding values of S will be found to be, after the manner described by Kendall (1946):

$$S = +70 - 9 + 21 = +82,$$

$$S = -30 + 49 - 21 = -2.$$

For the calculation of τ in the case of tied rankings we have a choice in the denominator by which S is to be divided to give τ. In the untied case this would be $\frac{1}{2}n(n-1)$, where n is

the number of ranks. In the tied case we may take the denominator S' as $\frac{1}{2}n(n-1)$ or as $[\frac{1}{2}n(n-1)\{\frac{1}{2}n(n-1)-\frac{1}{2}\Sigma t(t-1)\}]^{\frac{1}{2}}$, where t_1, t_2, etc., are the extent of the ties. The choice is determined by practical considerations (see Kendall, 1946), but is not material to a discussion of significance. For an untied ranking and a dichotomy with 'x' and 'y' members in each class, the second form reduces to $\{\frac{1}{2}xyn(n-1)\}^{\frac{1}{2}}$.

In the case of two untied rankings Kendall has shown that $\text{var}\, S = \frac{1}{18}n(n-1)(2n+5)$. In the case of one untied ranking and one with ties of extent t_1, t_2, etc., Sillitto (1947) has extended this result by proving that

$$\text{var}\, S = \tfrac{1}{18}\{n(n-1)(2n+5)-\Sigma t(t-1)(2t+5)\}. \qquad (1)$$

In the case of an untied ranking and a dichotomy, $t_1 = x$, $t_2 = y$, $(x+y) = n$, and we have then the simple form

$$\text{var}\, S = \tfrac{1}{3}xy(n+1). \qquad (2)$$

In the example above this gives

$$\text{var}\, S = \frac{(10)(10)(21)}{3} = 700, \quad \sqrt{(\text{var}\, S)} = 26{\cdot}46, \quad \frac{S_{AB}}{\sqrt{(\text{var}\, S)}} = \frac{82-1}{26{\cdot}46} = 3{\cdot}06.$$

The probability of a deviation greater than this in absolute value is 0·0022. Further,

$$\frac{S_{AC}}{\sqrt{(\text{var}\, S)}} = \frac{2-1}{26{\cdot}46} = 0{\cdot}0378,$$

and the corresponding probability is 0·970.

VARIANCE WHEN THERE ARE TIES IN THE RANKING

The variance of S given by equation (2) is true only in the case of a dichotomy and an untied ranking. For a tied ranking I surmised from some special cases that

$$\text{var}\, S = \frac{xy}{3n(n-1)}\{(n^3-n)-\Sigma(t^3-t)\}. \qquad (3)$$

In the note following this paper Mr Kendall provides proof of this result.

Example (from data collected by the Medical Research Council team in Germany 1946, as yet unpublished). Selected workers in a factory were interviewed and an assessment made of their adaptation to living conditions. They were assessed as 'Efficient' or 'Overactive'. Other data were available, including statements by the men of frequency of nocturia. For men aged 50–59 years the following was observed:

Assessment	Rank order of frequency of nocturia (least frequent nocturia given highest rank)
Efficient	2½, 2½, 2½, 2½, 6½, 6½, 10, 10, 10, 10, 14, 14
Overactive	5, 10, 14, 16, 17

Five is the highest ranking in the overactive group. Four members of the efficient group have higher rankings, and eight lower rankings. The S score for that member is therefore $4-8$. Similarly, for all members we have

$$S = 4-8+6-2+10+12+12 = +34.$$

Using a denominator in the form

$$[xy\{\tfrac{1}{2}n(n-1)-\tfrac{1}{2}\Sigma t(t-1)\}]^{\frac{1}{2}},$$

τ is given by

$$\tau = +\frac{34}{\sqrt{[(12)(5)\{\tfrac{1}{2}(17)(16)-\tfrac{1}{2}(4)(3)-\tfrac{1}{2}(2)(1)-\tfrac{1}{2}(5)(4)-\tfrac{1}{2}(3)(2)\}]}} = +\frac{34}{\sqrt{6960}} = +0{\cdot}408.$$

From (3) we then have

$$\text{var } S = \frac{(12)(5)}{3(17)(16)}\{(17^3-17)-(4^3-4)-(2^3-2)-(5^3-5)-(3^3-3)\}$$
$$= 344{\cdot}6.$$

A small problem arises when we consider the correction for continuity to be applied in testing the significance of an observed value of S. In the case of a dichotomy and an untied ranking the interval between successive S values is 2. In the case of a dichotomy and a ranking composed entirely of ties of the same extent 't', the interval is $2t$. But in the example the ties are of varying extent, and the interval between successive S values is composed of a mixture of the intervals produced by the successive rank values. Thus, although these varying intervals are combined so that over most of the range the interval between successive values is unity, the distribution oscillates somewhat, and to use the value $\frac{1}{2}$ as the correction for continuity would sometimes be misleading. Further work is required to determine the correction which will provide a probability on the normal distribution equal to or slightly greater than the true probability in all cases. Until this is available I propose to use a crude correction, based on the average of the intervals mentioned above. In the example the successive rank values $2\frac{1}{2}$ and 5 give an interval of 5 in S score, rank values 5 and $6\frac{1}{2}$ give an interval of 3, and it is therefore possible to determine the average interval by calculating the intervals given by successive rank values. This calculation can be shortened. The total of the S score intervals is twice the number of members, less the extent of the ties involving the first and last members. If we divide this by the number of intervals between successive rank values we have the average S score interval. In the example this is $\frac{1}{6}(34-4-1)$. Using half of this as the correction for continuity we have

$$\frac{S}{\sqrt{(\text{var } S)}} = \frac{34-2{\cdot}42}{18{\cdot}56} = 1{\cdot}702.$$

The pre-observational hypothesis, made on psychological grounds, was that excessive nocturia is a symptom of inefficient adaptation to living conditions, i.e. a positive correlation should be obtained. From these observations the probability of a positive correlation as great or greater than the observed value appearing by chance is 0·044. Direct calculation of the positive tail of the distribution of S gives a probability of 0·0368.

The alternative testing hypothesis based on the absolute value of S gives a probability twice as great, and the corresponding direct calculation using both positive and negative tails of the actual S distribution gives a probability of 0·0735.

By itself this evidence could only be debatable substantiation of the psychological hypothesis. In fact, additional data from two other factory groups, treated in the same way, gave a total S value of +104, the square-root of the total variance being 35·00, providing a justification of the hypothesis.

THE CASE OF THE 2×2 TABLE

If one dichotomous variable can be considered as a ranking with two sets of tied ranks it is logical to consider the case when both variables are in this form. If we have a 2×2 table in the form

(AB)	(Ab)	(A)
(aB)	(ab)	(a)
(B)	(b)	N

any member of (AB) taken with any member of (ab) has the same order in either ranking and hence contributes $+1$ to S, and any member of (Ab) with any member of (aB) contributes -1. The others contribute nothing. Hence

$$S = (AB)(ab) - (Ab)(aB).$$

From equation (3)

$$\text{var } S = \frac{(A)(a)}{3N(N-1)}[(N^3 - N) - \{(B)^3 - (B)\} - \{(b)^3 - (b)\}]$$

$$= \frac{(A)(a)(B)(b)}{N-1}. \tag{4}$$

Again, for testing the significance of an observed value of S it is necessary to correct for continuity by subtracting half the interval between successive S values. In the case of the 2×2 table the interval is N, for if we increase (AB) by unity S becomes

$$\{(AB)+1\}\{(ab)+1\} - \{(Ab)-1\}\{(aB)-1\} = (AB)(ab) - (Ab)(aB) + N.$$

Hence, for the normal deviate, we have

$$\frac{S - \frac{1}{2}N}{\sqrt{\left\{\frac{(A)(a)(B)(b)}{N-1}\right\}}}. \tag{5}$$

It will be noted that τ (taking the ties into account in calculating the denominator S') is

$$\frac{(AB)(ab) - (Ab)(aB)}{\sqrt{[(A)(a)(B)(b)]}},$$

which is the product-moment correlation for a 2×2 table when the variables are conventionally regarded as possessing the discrete values 0, 1.

Testing by use of the normal deviate seems to be moderately accurate, and would appear to be useful in those cases where χ^2 is suspect because of small expectations in the cells of the 2×2 table. It is less laborious to calculate than the hypergeometric treatment, and is an alternative form of the approximation to hypergeometric treatment given by Pearson (1947), who also discusses the order of accuracy of the approximation.

Using the data given earlier as an example, but assuming that it had been possible only to grade nocturia into 'Normal' or 'Excessive', we have the following table:

Assessment	Nocturia	
	Normal	Excessive
Efficient	10	2
Overactive	2	3

$$S = 30-4 = 26, \quad \text{var}\, S = \frac{(12)(5)(12)(5)}{16} = 225.$$

This gives, after correction for continuity,

$$\frac{S-\frac{1}{2}N}{\sqrt{(\text{var}\, S)}} = \frac{26-8{\cdot}5}{15} = 1{\cdot}1667.$$

This gives the probability of S being attained or exceeded in the direction of the hypothesis (i.e. positive values only) as 0·1217. χ^2 without the continuity correction gives $P = 0{\cdot}0369$,* and with the correction, $P = 0{\cdot}1143$. The hypergeometric treatment, summing the probabilities of obtaining 3, 4 or 5 in the Overactive-Excessive category, gives $P = 0{\cdot}1166$.

If the more customary test of absolute value is applied, χ^2 with Yates's correction gives $P = 0{\cdot}2286$, S and the normal deviate gives $P = 0{\cdot}2434$, i.e. both values of P are doubled. The hypergeometric treatment, adding the probability of obtaining 0 in the Overactive-Excessive category gives $P = 0{\cdot}2445$.

It will be seen that in conditions such as these, S and the normal deviate give a reasonable approximation to the exact treatment.

* This is making the common assumption that $\{(AB)(ab)-(Ab)(aB)\}^2 N/\{(A)(a)(B)(b)\}$ is distributed as χ^2 with 1 degree of freedom, or that its square root is a normal deviate with sign depending on the sign of $(AB)(ab)-(Ab)(aB)$.

REFERENCES

Chambers, E. G. (1946). Statistical techniques in applied psychology. *Biometrika*, **33**, 269.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

Kendall, M. G. (1946). The treatment of ties in ranking problems. *Biometrika*, **33**, 239.

Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139.

Sillitto, G. P. (1947). The distribution of Kendall's coefficient of rank correlation in rankings containing ties. *Biometrika*, **34**, 36.

THE VARIANCE OF τ WHEN BOTH RANKINGS CONTAIN TIES

By M. G. KENDALL

1. The variance of τ in the population of sample permutations was given in my paper of 1938 for the case where no tied ranks exist. Mr Sillitto (1947) has given the formula where one ranking contains ties but the other does not. In the foregoing paper Mr Whitfield has correctly surmised the variance when one ranking contains ties, and the other is a dichotomy. In this note I derive the general formula for the variance when both rankings contain ties. The results of Messrs Sillitto and Whitfield then follow as special cases.

2. I shall follow the method of Daniels (1944). If a_{ij} represents the contribution of the ith and jth members of a ranking to τ we have

$$\begin{aligned} a_{ij} &= +1 \quad (i<j) \\ &= \ 0 \quad (i=j) \\ &= -1 \quad (i>j) \end{aligned} \tag{1}$$

We write

$$c_{ij} = a_{ij} b_{ij}, \tag{2}$$

where a and b refer to different rankings, and

$$c = \sum_{i,j=1}^{n} c_{ij}. \tag{3}$$

The quantity c is simply related to S by the relation

$$c = 2S, \tag{4}$$

and for the testing of τ it is sufficient to test c or S which are merely constant multiples of τ. I work with the quantity c.

3. We have, from Daniel's results,

$$\sum_{l=1}^{n} a_{il} = n+1-2i, \tag{5}$$

$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1), \tag{6}$$

$$\sum_{i,l,l=1}^{n} a_{il}\, a_{il} = \tfrac{1}{3}n(n^2-1), \tag{7}$$

$$E(c) = 0, \tag{8}$$

$$E(c^2) = \frac{4}{n(n-1)(n-2)}\{\Sigma a_{il} a_{il} - \Sigma a_{il}^2\}\{\Sigma b_{il} b_{il} - \Sigma b_{il}^2\} + \frac{2}{n(n-1)}\Sigma a_{il}^2 \Sigma b_{il}^2. \tag{9}$$

If we substitute from (6) and (7) in (9) we find

$$E(c^2) = \tfrac{2}{9}n(n-1)(2n+5), \tag{10}$$

or, equivalently,

$$E(S^2) = \tfrac{1}{18}n(n-1)(2n+5), \tag{11}$$

from which the variance of τ in the case of untied rankings follows at once.

4. Now suppose that sets of t_1, t_2, ... consecutive members in one ranking are tied. In place of (6) we then have

$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1) - \Sigma t(t-1), \tag{12}$$

the summation on the right taking place over the various values of t. This result follows simply from the consideration that for a pair of tied ranks $a_{ij} = 0$, and consequently the sum of squares of contributions from a tied set is of the same form as for the ranking as a whole.

In place of (7) we have

$$\sum_{i,l,j=1}^{n} a_{il}a_{jl} = \tfrac{1}{3}n(n^2-1) - \tfrac{1}{3}\Sigma t(t^2-1). \tag{13}$$

This is not quite so obvious. Consider a set of tied ranks. The contribution to the sum on the left of (13) will be unchanged if the suffixes l, t fall outside this set. If they both fall inside, no contribution arises and therefore we have to subtract the term $\frac{1}{3}t(t^2-1)$. The remaining possibility is that one falls inside and one outside. In such a case the contribution remains unchanged in total for it is zero in the original untied case, each possible pair occurring once to give $+1$ and one to give -1. Formula (13) follows.

5. By substitution in (9) we then have, for two rankings with ties typified respectively by t and u,

$$\begin{aligned} E(c^2) = {} & \frac{4}{n(n-1)(n-2)}\{\tfrac{1}{3}n(n-1)(n-2) - \tfrac{1}{3}\Sigma t(t-1)(t-2)\} \\ & \times \{\tfrac{1}{3}n(n-1)(n-2) - \tfrac{1}{3}\Sigma u(u-1)(u-2)\} \\ & + \frac{2}{n(n-1)}\{n(n-1) - \Sigma t(t-1)\}\{n(n-1) - \Sigma u(u-1)\}. \end{aligned} \tag{14}$$

This is the general formula required. We can express it in the alternative form

$$\begin{aligned} E(c^2) = {} & \tfrac{2}{9}n(n-1)(2n+5) - \tfrac{2}{9}\Sigma t(t-1)(2t+5) - \tfrac{2}{9}\Sigma u(u-1)(2u+5) \\ & + \frac{4}{9n(n-1)(n-2)}\{\Sigma t(t-1)(t-2)\}\{\Sigma u(u-1)(u-2)\} \\ & + \frac{2}{n(n-1)}\{\Sigma t(t-1)\}\{\Sigma u(u-1)\}. \end{aligned} \tag{15}$$

6. (i) If one ranking is untied, say all the u's are zero, we have Mr Sillitto's result

$$E(c^2) = \tfrac{2}{9}n(n-1)(2n+5) - \tfrac{2}{9}\Sigma t(t-1)(2t+5). \tag{16}$$

(ii) If one ranking is untied and the other is a dichotomy into x and $n-x=y$ members, (16) reduces to

$$E(c^2) = \tfrac{4}{3}xy(n+1), \tag{17}$$

agreeing with Mr Whitfield's equation (2).

(iii) If one ranking contains ties and the other is a dichotomy we find on substitution in (14)

$$E(c^2) = \frac{4xy}{3n(n-1)}\{n^3 - n - \Sigma(t^3-t)\}, \tag{18}$$

agreeing with Mr Whitfield's equation (3).

(iv) Finally, if both variates are dichotomized into x, y and p, q we find

$$E(c^2) = \frac{4xypq}{n-1}, \tag{19}$$

agreeing with Mr Whitfield's equation (4).

REFERENCES

See the references to Mr Whitfield's paper together with:

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129.

A χ^2 'SMOOTH' TEST FOR GOODNESS OF FIT

By F. N. DAVID

1. The χ^2 test occupies a central position in statistical theory, and it is difficult to imagine another test which would have the same generality of application. We shall be concerned here with one aspect only, that is, the uses of χ^2 in the tests for agreement between hypothesis and observation which are usually loosely classed together under the name of tests for goodness of fit. The principal advantages of χ^2 for such tests would seem to be (i) that it is applicable to grouped observations, (ii) that the parameters of the hypothesis tested may be calculated from the observational data and the fact allowed for in the degrees of freedom with which the criterion is assumed to be distributed and (iii) that it is easy to calculate, for the number of computations involved is just the number of groups into which the observational data are divided. It has, however, two defects which have long been known and which are easily recognized from the form of the criterion itself. Broadly speaking we may define χ^2 as follows:

$$\chi^2 = \text{sum for all groups of } \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}.$$

It will be seen (iv) that in taking the square of (Observed value − Expected value) the knowledge regarding the sign of the deviation is lost. Further (v) there is no means of preserving the order of the signs of the deviations, and no distinction can therefore be made between a departure from the hypothesis tested in which the deviations were first all positive and then all negative and a departure in which the sizes and signs of the deviations were random between themselves.

2. The ideal test for goodness of fit should certainly take into account (i), (ii), (iv) and (v) and probably (iii) also, but it is unlikely that this ideal will be reached. It would seem at the present time that the most which may be hoped for are tests which will supplement the χ^2 test in that they will be more sensitive to given alternatives of the hypothesis under test. Neyman (1937) put forward such a supplementary test in which he developed the ψ^2 criterion. This criterion was designed to be sensitive to alternate hypotheses of a type he designated as smooth; that is to say, if the hypothesis under test is a continuous curve, such as, for example, the normal curve, admissible hypotheses alternate to it might be other normal curves with a different mean and a different standard deviation. Neyman's criterion certainly took into account (iv) and (v), but it did not fulfil (i), in that it was only applicable to ungrouped observations,* or (ii), in that the parameters of the hypothesis tested had to be known *a priori*. Whether it fulfilled (iii) is a matter for personal opinion.

3. It would appear possible that tests which would take into account the sign only of the deviations of observation from hypothesis, and the order of these signs, may be devised from simple combinatorial principles. Suppose that there are N observations which are divided into k groups. Let n_i $(i = 1, 2, \ldots, k)$ be the actual number of observations falling

* Prof. Neyman tells me that his criterion has been adapted for the case of grouped observations but that he has not yet published this extension.

into the ith group, and let m_i ($i = 1, 2, \ldots, k$) be the expected number. It is possible theoretically for χ^2 to be calculated for the case where

$$\sum_{i=1}^{k} n_i = N \neq \sum_{i=1}^{k} m_i,$$

but such cases must be rare in statistical practice. We shall overlook this case and will consider the case where the totals of observed and expected are made equal to one another with the resultant loss of one degree of freedom in the calculation of χ^2. If the totals agree then

$$\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} m_i = \sum_{i=1}^{k} \delta_i = 0,$$

where $\delta_i = n_i - m_i$. In order that the sum of these δ's should be zero, at least one of them must be negative in sign, but which one of these δ's it will be would seem to be a matter of chance. It is on this fact that we shall base the first test criterion.

4. Suppose that we have a sequence of signs of which r_1 are positive and r_2 negative, where $r_1 + r_2 = r$, and $r_1 > 0$ and $r_2 > 0$. These signs are postulated to occur in a random order. Given such a sequence it is easy to record the number of sets of positive and negative signs. For example, if the sequence is

$$+ + - - + + - - + - + + + + -,$$

then $r = 15$, $r_1 = 9$, $r_2 = 6$, and there are four sets of positive signs and four sets of negative signs. In general there can be (α) t positive, t negative, or (β) t positive, $t+1$ negative, or (γ) $t+1$ positive and t negative sets of signs. If $T = 2t$ or $2t+1$ as required, we may ask what is the probability that given r_1 and r_2 such a number T of sets (alternately positive and negative) would have arisen through chance. This probability follows at once from Whitworth, *Choice and Chance*, Proposition XXV, viz.: 'The number of ways in which n indifferent things can be distributed in t different parcels (blank lots being inadmissible) is

$$(n-1)!/(t-1)!\,(n-t)!'.*$$

5. The total number of ways in which r_1 and r_2 elements can be arranged is

$$\frac{r!}{r_1!\,r_2!}.$$

We now require to enumerate the number of ways in which r_1 can be arranged to form t sets and r_2 to form t sets. To arrange r_1 in t sets is equivalent (vide Whitworth) to making $t-1$ breaks in a sequence of r_1 observations, and this may be done in

$$(r_1-1)!/(t-1)!\,(r_1-t)! \text{ ways},$$

and similarly for r_2. It is not specified whether $+$ or $-$ should start the sequence, and hence the total number of ways in which a sequence $r_1 + r_2$ may be arranged in t sets each is

$$\frac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}.$$

* Since I first thought of this method of attack I have found that the distribution of groups as given by me in § 5 has already been given by W. L. Stevens, *Ann. Eugen., Lond.*, **9**, 10, and by A. Wald and J. Wolfowitz, *Ann. Math. Statist.* **11**, 147. The probability function has been tabled by F. S. Smed and C. Eisenhart, *Ann. Math. Statist.* **14**, 66, but it is not in a form that I found suitable for my purposes. The probability function has been known for many years; what is interesting is the different uses to which it has been put.

The probability of $2t$ sets will be

$$P\{2t \mid r_1, r_2\} = \frac{2r_1!(r_1-1)!\,r_2!(r_2-1)!}{r!(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}, \tag{1}$$

and the probability of obtaining $(2t+1)$ sets will be

$$P\{2t+1 \mid r_1, r_2\} = P\{t \mid r_1, t+1 \mid r_2\} + P\{t+1 \mid r_1, t \mid r_2\} = P\{2t \mid r_1, r_2\}\left(\frac{r-2t}{2t}\right). \tag{2}$$

Hence given r_1, r_2 and T from a random sequence of positive and negative signs the probability of such a number of sets having arisen through chance may be calculated.

6. It is desired to use the probability of a given arrangement of signs in order to test a given hypothesis represented by a smooth probability law, bearing in mind that, if the given hypothesis is not true, then any alternative law is likely to be of a smooth type. Although no exact definition of a smooth alternative distribution has been made, it may be stated here that *smooth*, in the sense used by Neyman, will imply that the number of sets of signs will be small. For example, if the hypothesis tested is that observations follow a given normal curve, whereas in fact they have been drawn from a normal distribution identical with the first but with a smaller mean, then the differences between observation and expectation on the basis of the hypothesis tested may be expected to give a preponderance of positive signs below the sample mean and of negative signs above it; that is to say, if the difference in means is sufficient to offset the sampling fluctuations we should find a single set of positive signs followed by a single set of negative signs. If the true population is a normal curve with the same mean but with a larger standard deviation than that specified by the hypothesis tested, then there will be a tendency towards a set of positive signs, a set of negative signs, followed by a set of positive signs, although sampling fluctuations may not leave such a clear-cut answer. The more complex the alternative hypothesis the less chance there will be of detecting it.

7. With this objective in view it is proposed to take T, the number of sets of signs, as the test criterion, rejecting the hypothesis tested whenever, for a given r_1 and r_2, T is exceptionally small. This we do on the grounds that the existence of very few sets of signs suggests that the differences between observed and expected frequencies are not due to chance sampling fluctuations but to some systematic departure of the true probability law (assumed *smooth*) from hypothesis. In following this procedure we should reject the hypothesis if

$$P\{T \leqslant T_0\} = \sum_{T=2}^{T_0} P\{T \mid r_1 r_2\} \leqslant \epsilon,$$

where T_0 is the observed value of T and ϵ the significance level selected as appropriate. Exact probabilities are given in Table 1, and the application of the test is immediate.* There seems to be no reason why the test should not be applicable to both grouped and ungrouped observations, although the formulation of the hypothesis tested may be somewhat different in the two cases. Consider a sample which has been supposedly randomly

* An assumption implicit in the test would appear to be that for each χ^2 cell there is an equal chance of obtaining a positive or a negative deviation, that is, that there are sufficient numbers in each cell for the binomial to be closely approximated to by a normal curve. An extensive series of random sampling experiments has shown, however, that the divergence between theory and practice is not significant even when the probability of obtaining a positive is four times that of obtaining a negative. Hence while strictly the expectation in each cell of χ^2 should be 10 or over, it would seem that for practical purposes that the T test may be applied in all cases where the application of the χ^2 test is permissible.

Table 1. *Probability of obtaining a given number of sets, T.* $[T = 2t \text{ or } 2t+1]$

The function tabled is $\dfrac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}$ or $\dfrac{(r_1-1)!\,(r_2-1)!\,(r_1+r_2-2t)}{t!(t-1)!\,(r_1-t)!\,(r_2-t)!}$, according as T is even or odd.

$P\{T\}$ is obtained by dividing this function by the binomial term $\dfrac{r!}{r_1!\,r_2!}$.

r	r_1	r_2	$\frac{r!}{r_1!\,r_2!}$	$T=2$	3	4	5	6	7	8	9	10	11	12	13	14
2	1	1	2	2												
3	2	1	3	2	1											
4	3	1	4	2	2											
	2	2	6	2	2	2										
5	4	1	5	2	3											
	3	2	10	2	3	4	1									
6	5	1	6	2	4											
	4	2	15	2	4	6	3									
	3	3	20	2	4	8	4	2								
7	6	1	7	2	5											
	5	2	21	2	5	8	6									
	4	3	35	2	5	12	9	6	1							
8	7	1	8	2	6											
	6	2	28	2	6	10	10									
	5	3	56	2	6	16	16	12	4							
	4	4	70	2	6	18	18	18	6	2						
9	8	1	9	2	7											
	7	2	36	2	7	12	15									
	6	3	84	2	7	20	25	20	10							
	5	4	126	2	7	24	30	36	18	8	1					
10	9	1	10	2	8											
	8	2	45	2	8	14	21									
	7	3	120	2	8	24	36	30	20							
	6	4	210	2	8	30	45	60	40	20	5					
	5	5	252	2	8	32	48	72	48	32	8	2				
11	10	1	11	2	9											
	9	2	55	2	9	16	28									
	8	3	165	2	9	28	49	42	35							
	7	4	330	2	9	36	63	90	75	40	15					
	6	5	462	2	9	40	70	120	100	80	30	10	1			
12	11	1	12	2	10											
	10	2	66	2	10	18	36									
	9	3	220	2	10	32	64	56	56							
	8	4	495	2	10	42	84	126	126	70	35					
	7	5	792	2	10	48	96	180	180	160	80	30	6			
	6	6	924	2	10	50	100	200	200	200	100	50	10	2		
13	12	1	13	2	11											
	11	2	78	2	11	20	45									
	10	3	286	2	11	36	81	72	84							
	9	4	715	2	11	48	108	168	196	112	70					
	8	5	1287	2	11	56	126	252	294	280	175	70	21			
	7	6	1716	2	11	60	135	300	350	400	250	150	45	12	1	
14	13	1	14	2	12											
	12	2	91	2	12	22	55									
	11	3	364	2	12	40	100	90	120							
	10	4	1001	2	12	54	135	216	288	168	126					
	9	5	2002	2	12	64	160	336	448	448	336	140	56			
	8	6	3003	2	12	70	175	420	560	700	525	350	140	42	7	
	7	7	3432	2	12	72	180	450	600	800	600	450	180	72	12	2

drawn from some population. Let the elements of the sample in order of drawing be $x_1, x_2, \ldots, x_n$. We may use the T criterion to test the hypothesis of randomness, in the following way. If u_1 is the smallest value observed in the sample and u_n the largest, then if we exclude the trivial case when all the x's are equal it is easy to show that $u_1 < \bar{x} < u_n$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. If we now consider the deviations

$$x_i - \bar{x} = \delta x_i \quad \text{for} \quad i = 1, 2, \ldots, n,$$

there will be a series $\delta x_1, \delta x_2, \ldots, \delta x_n$, some of which quantities will be positive and some negative. The application of the T test is immediate, the admissible alternate hypotheses being that if the drawing of the sample is not at random then bias of the smooth kind is present.

8. As an illustration consider the following two cases:

Case I.	Expected frequency	10	25	35	75	155	155	75	35	25	10
	Observation	12	29	45	81	160	145	69	31	20	8
	Deviation	+	+	+	+	+	−	−	−	−	−
Case II.	Expected frequency	10	25	35	75	155	155	75	35	25	10
	Observation	12	23	45	66	161	160	69	36	20	8
	Deviation	+	−	+	−	+	+	−	+	−	−

In the first case $\chi^2 = 6{\cdot}94$ and in the second $\chi^2 = 6{\cdot}80$; in neither case would the hypothesis be rejected as inadequate by using the χ^2 criterion. The T criterion does, however, bring out the essential difference:

Case I. $r_1 = 5, \quad r_2 = 5, \quad T_0 = 2 \quad$ and $\quad P\{T \leqslant T_0\} = \frac{2}{252}$.

Case II. $r_1 = 5, \quad r_2 = 5, \quad T_0 = 8 \quad$ and $\quad P\{T \leqslant T_0\} = \frac{242}{252}$.

Using the T test we should be inclined to reject the first hypothesis in favour of a smooth alternative, while for the second case we should be inclined to agree with the conclusion drawn from the χ^2 test that the observational material is adequately described.

9. Sampling material is available whereby the theoretical distribution of T may be tested in practice. Neyman & Pearson (1928) took 208 samples, each of size 200, from a population of eight groups described by the cubic curve

$$y = 25 + \tfrac{45}{8}x - \tfrac{15}{128}x^3.$$

The expectation in each cell for a sample of this size was calculated and the χ^2 criterion found for each of the 208 samples. The writer was given access to these calculations and was able to find the sampling distribution of T from the material. The results of this sampling experiment and the theoretical distribution of T from relations (1) and (2) are given in Table 2.

The agreement between theory and practice would seem to be reasonably good, and in the cases (4, 4) and (5, 3) the values of χ^2, calculated to test the discrepancy between theory and practice, were not greater than might be attributable to sampling fluctuations. It was not thought worth while to calculate χ^2 for (6, 2) and (7, 1). A second sampling experiment in which samples of size 360 were drawn from a normal population of fifteen groups lent further support to the reasonableness of the theoretical distribution.

10. The T criterion will be a useful supplementary criterion to the χ^2, but because it takes account solely of the sign of a distribution and not of its magnitude it will probably only be useful when used in conjunction with χ^2. A test of significance which could combine both the probability levels of T and χ^2 would undoubtedly be more useful, and we may

therefore consider how this might be done. Unless the exact degree of dependence which exists between two variables is known it is usually only possible to obtain their joint distribution if they are independent. It would appear reasonable, both on theoretical grounds and from sampling experiments, to assume that T and χ^2 are independent, or, if the assumptions underlying both tests are not exactly fulfilled, to assume that the degree of dependence between them is at most small.

Table 2. *Comparison of theoretical distribution of T with that derived from a sampling experiment*

(4 positive, 4 negative)

T = number of sets	2	3	4	5	6	7	8	Total
Sampling	3	5	20	25	28	5	7	93
Theory	2·7	8·0	23·9	23·9	23·9	8·0	2·7	93

(5 positive, 3 negative) or (3 positive, 5 negative)

T = number of sets	2	3	4	5	6	7	Total
Sampling	2	8	32	30	20	10	102
Theory	3·6	10·9	29·1	29·1	21·9	7·3	102

(6 positive, 2 negative) or (2 positive, 6 negative)

T = number of sets	2	3	4	5	Total
Sampling	1	3	—	5	9
Theory	0·6	1·9	3·2	3·2	9

(7 positive, 1 negative) or (1 positive, 7 negative)

T = number of sets	2	3	Total
Sampling	—	2	2
Theory	0·5	1·5	2

11. We shall begin by demonstrating that as far as mathematics are concerned the T and χ^2 criteria are completely independent.* For simplicity of argument let us consider the case of three groups only. The sample may then be represented by a point (n_1, n_2, n_3) in three-dimensioned space, with axes of reference On_1, On_2, On_3, and the expected population values by a point (m_1, m_2, m_3) in the same space. Since

$$n_1+n_2+n_3 = m_1+m_2+m_3 = N,$$

* This method of approach was suggested to me by Andrew Gleason of Harvard University at a seminar given at the Statistical Laboratory, University of California, at Berkeley.

these points are constrained to lie in a plane. Fig. 1 shows this plane for the particular case $N = 16; m_1 = 4, m_2 = 8, m_3 = 4$. Since no frequency can be negative, possible sample points must be within an equilateral triangle lying in this plane, the chance of occurrence associated with a point being the multinomial term

$$\frac{N!}{n_1!\,n_2!\,n_3!}\left(\frac{m_1}{N}\right)^{n_1}\left(\frac{m_2}{N}\right)^{n_2}\left(\frac{m_3}{N}\right)^{n_3}.$$

When using the χ^2 test the mathematical approximation consists in substituting for this term an expression proportional to $e^{-\frac{1}{2}\chi^2}$, in regarding this last as a continuous function, and

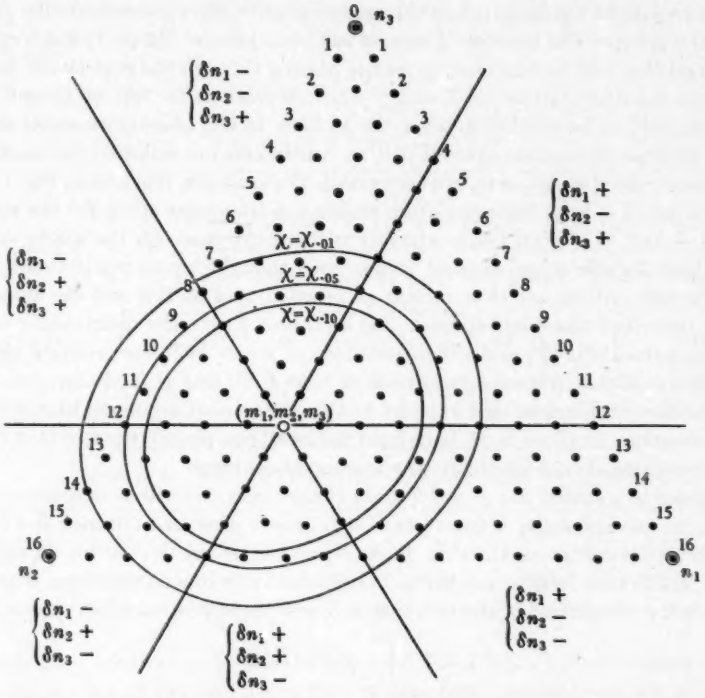

0 n_3
$\delta n_1 -$ $\delta n_2 -$ $\delta n_3 +$
$\delta n_1 +$ $\delta n_2 -$ $\delta n_3 +$
$\delta n_1 -$ $\delta n_2 +$ $\delta n_3 +$
$\chi=\chi_{.01}$
$\chi=\chi_{.05}$
$\chi=\chi_{.10}$
(m_1, m_2, m_3)
$\delta n_1 -$ $\delta n_2 +$ $\delta n_3 -$
$\delta n_i +$ $\delta n_2 +$ $\delta n_3 -$
$\delta n_1 +$ $\delta n_2 -$ $\delta n_3 -$
n_2
n_1

Fig. 1. Graphical illustration of the χ^2 contours and the change in signs of the δn's. n_1, n_2 and n_3 denote the points of intersection of the On_1, On_2, On_3 axes with the plane $n_1+n_2+n_3 = N$. According to the approximation, the chance equals α of obtaining a sample point lying outside the elliptic contour on which $\chi = \chi_\alpha$.

in taking as a measure of goodness of fit the integral of this expression outside the ellipse which passes through the sample point and on which χ^2 is constant. For the case of three groups this integral itself assumes the simple form $e^{-\frac{1}{2}\chi^2}$. Three such elliptic contours are shown in the diagram.

Planes through (m_1, m_2, m_3) parallel to the co-ordinate planes n_1On_2, n_2On_3, n_3On_1, will intersect the sample plane

$$n_1+n_2+n_3 = N$$

in three straight lines. As shown in the diagram, these lines divide the sample plane into six sectors, and for all sample points within a sector the signs of the differences $\delta n_i = n_i - m_i$ will remain unchanged. Any test based solely on runs of signs will consist in taking one or more of these sectors as critical regions and rejecting the hypothesis tested when the sample point falls therein. It is clear that if we use the mathematical approximation, the distribution of χ^2 is the same within each sector; similarly, that the chance of a sample showing a given combination of signs is the same on each ellipse along which χ^2 is constant. Thus under the assumptions made regarding the distribution of χ^2, the T and χ^2 criteria are completely independent.

In this case of three groups T can only assume values of two or three and the former value would not be judged significant, but the argument will follow exactly similar lines in the case of many groups. The number of sectors will be in general $2(2r_2+1)$ if $r_1 > r_2$ and $2(2r_2)$ if $r_1 = r_2$, and they will be bounded by primes passing through the population point.

12. While the distributions of T and χ^2 are independent for this mathematical model they are unlikely to be exactly so when we go back to the true multinomial density distribution, because the sample space is neither continuous nor infinite. The model, in fact, becomes inaccurate if m_1, m_2 or m_3 are very small. For example, it is seen in Fig. 1 that while the 1 % ellipse ($\chi = \chi_{0\cdot01}$) lies completely within the triangular space for the sectors with signs $+\ -\ +$ and $-\ -\ +$, it lies completely without the space for the sector $+\ +\ -$ and partly without for the other sectors. It has been thought worth while therefore to test whether the two criteria are independent in practice, and to this end the same material previously described has been utilized. Tables 3 and 4 give the distribution of mean χ^2 for different values of $P\{T\}$ and the distribution of mean $P\{T\}$ for grouped values of χ^2. There is little evidence in these figures to show that $P\{T\}$ and χ^2 (and therefore $P\{\chi^2\}$) are related. The figures therefore lend support to the geometrical argument and indicate that the approximations involved in χ^2, both from the small sample and the fact that the sample space is not infinite, do not invalidate the mathematical result.

13. In order to combine the χ^2 and T tests of significance it will be necessary to develop a theory for the combination of two tests of significance when one criterion is a continuous and the other a discontinuous variable. R. A. Fisher has set out the test for the combination of tests of significance from a number of independent continuous variables. The keystone of the test is the recognition of the fact that if Z is a continuous variable, then z, where

$$z = \int_{-\infty}^{Z} p(Z)\,dZ,$$

is also a continuous variable equally likely to have any value between 0 and 1; we shall describe z as being distributed rectangularly. Twice the logarithm of the product of two such z's, say z_1 and z_2, where z_1 and z_2 follow from two independent tests of significance can be shown to be distributed as χ^2 with four degrees of freedom. Consider a discontinuous variable X which may take values $X_1, X_2, \ldots, X_s$ and which has an elementary probability law $P\{X = X_j\} = p_j$, where $0 < p_j < 1$ for $j = 1, 2, \ldots, s$

and

$$\sum_{j=1}^{s} p_j = 1.$$

If a new variable, x, is defined as taking values $x_1, x_2, \ldots, x_s$, where

$$x_k = \sum_{j=1}^{k} p_j,$$

Table 3. *Mean χ^2 for different values of $P\{T\}$*

r_1, r_2 or r_2, r_1	—	4, 4	5, 3	4, 4	5, 3	6, 2	4, 4	5, 3	4, 4	6, 2
No. of obs. on which mean is based	26	5	20	28	30	—	25	32	20	3
$P\{T\}$	1·00	0·97	0·93	0·86	0·71	0·64	0·63	0·43	0·37	0·29
Mean χ^2	7·19	6·04	5·87	6·43	5·98	—	7·41	7·32	6·82	4·86

r_1, r_2 or r_2, r_1	7,1	5, 3	4, 4	6, 2	5, 3 and 4, 4
No. of obs. on which mean is based	—	8	5	1	5
$P\{T\}$	0·25	0·14	0·11	0·07	0·04 and 0·03
Mean χ^2	—	6·10	5·53	6·95	8·06

Table 4. *Mean $P\{T\}$ for grouped χ^2*

χ^2	0·0–1·0	1·0–2·0	2·0–3·0	3·0–4·0	4·0–5·0	5·0–6·0	6·0–7·0	7·0–8·0	8·0–9·0	9·0–10·0
No. of obs. on which mean is based	1	7	14	27	32	29	21	14	11	15
Mean $P\{T\}$	0·71	0·73	0·69	0·60	0·68	0·60	0·65	0·57	0·67	0·64

χ^2	10·0–11·0	11·0–12·0	12·0–13·0	13·0–14·0	14·0–15·0	15·0–16·0	16·0–17·0	17·0–18·0
No. of obs. on which mean is based	10	10	5	4	2	1	2	1
Mean $P\{T\}$	0·74	0·65	0·74	0·58	0·82	1·00	0·63	0·63

then x_k may only take values between 0 and 1 for $k = 1, 2, \ldots, s$. It is required to find the joint probability law of the product of two independent variables x and z, where x and z are as defined above. It will be noted that the elementary probability law of x will be

$$P\{x = x_j\} = p_j \quad (j = 1, 2, \ldots, s).$$

Hence when $x = x_j$ (the probability of which is p_j), the product xz will be distributed rectangularly between 0 and x_j on a proportion p_j of occasions. It follows that xz has a probability distribution which has points of discontinuity at $x_1, x_2, \ldots, x_s$, that it is distributed rectangularly between these points of discontinuity, and that

$$P\{0 < xz < x_1\} = p_1 \sum_{j=1}^{s} \frac{p_j}{x_j}, \quad P\{x_1 < xz < x_2\} = p_2 \sum_{j=2}^{s} \frac{p_j}{x_j}.$$

Generally

$$P\{x_{k-1} < xz < x_k\} = p_k \sum_{j=k}^{s} \frac{p_j}{x_j}.$$

14. If we now apply this theory to the combination of the tests of significance of T and χ^2, it is seen that we must consider the product of $P\{\chi^2\}$ and $P\{T\}$. χ^2 is a continuous variable and

$$z = \int_{\chi_0^2}^{+\infty} p(\chi^2)\, d(\chi^2) = P\{\chi^2 \geqslant \chi_0^2\} = P\{\chi^2\}$$

is distributed rectangularly between 0 and 1, and

$$x = \sum_{T=2}^{T_0} P\{T \mid r_1 r_2\} = P\{T \leqslant T_0\} = P\{T\}$$

is a discontinuous variable taking known values. The probability integral of xz is thus known from theory and $Y_{0\cdot05}$ or $Y_{0\cdot01}$ can be found to satisfy the relation

$$P\{0 < xz < Y_\epsilon\} = \epsilon.$$

These probability levels are given in Table 5. The procedure for the joint test of significance will be:

(i) calculate $P\{T\}$ as described in § 7;

(ii) calculate $P\{\chi^2\}$ in the usual way. The degrees of freedom will be the number of groups minus one;

(iii) multiply $P\{T\}$ and $P\{\chi^2\}$ together and refer to Table 5 to judge the significance of the product.

Table 5. *Values of $Y_{0\cdot05}$ and $Y_{0\cdot01}$, where $P\{P(\chi^2)\,P(T) < Y_\epsilon\} = \epsilon$*

This table may be used to judge the significance of the joint distribution of the T criterion and any other continuous criterion.

r	r_1	r_2	$Y_{0\cdot05}$	$Y_{0\cdot01}$	r	r_1	r_2	$Y_{0\cdot05}$	$Y_{0\cdot01}$
5	4	1	0·0312^5	0·0062^5	11	10	1	0·0275$^+$	0·0055$^+$
	3	2	0·0213	0·0043		9	2	0·0171	0·0034
						8	3	0·0144	0·0028
6	5	1	0·0300	0·0060		7	4	0·0144	0·0025$^+$
	4	2	0·0211	0·0042		6	5	0·0140	0·0024
	3	3	0·0195	0·0039					
					12	11	1	0·0273	0·0055$^-$
7	6	1	0·0292	0·0058		10	2	0·0174	0·0035$^-$
	5	2	0·0197	0·0039^5		9	3	0·0149	0·0027
	4	3	0·0174	0·0035		8	4	0·0142	0·0024
						7	5	0·0135	0·0022
8	7	1	0·0286	0·0057		6	6	0·0131	0·0021
	6	2	0·0188	0·0038					
	5	3	0·0160	0·0032	13	12	1	0·0271	0·0054
	4	4	0·0153	0·0031		11	2	0·0165$^+$	0·0033
						10	3	0·0151	0·0026
9	8	1	0·0281	0·0056		9	4	0·0138	0·0023
	7	2	0·0180	0·0036		8	5	0·0137	0·0022
	6	3	0·0153	0·0031		7	6	0·0138	0·0022^5
	5	4	0·0140	0·0028					
					14	13	1	0·0269	0·0054
10	9	1	0·0278	0·0056		12	2	0·0163	0·0033
	8	2	0·0175	0·0035		11	3	0·0151	0·0025$^+$
	7	3	0·0143	0·0029		10	4	0·0135$^-$	0·0022^5
	6	4	0·0143	0·0026		9	5	0·0138	0·0023
	5	5	0·0143	0·0025		8	6	0·0136	0·0022
						7	7	0·0134	0·0022

15. The application of the joint test of significance may be illustrated by means of an example. A sample of 360 observations is available. This sample has actually been randomly drawn from a normal population of which the mean is zero and the standard deviation unity. The figures are given in Table 6. Calculations give $\chi^2 = 21{\cdot}1$ and $P\{\chi^2\} = 0{\cdot}10$. Judging by the χ^2 alone we should say probably that there is nothing out of the ordinary in the deviations of the sample from the expected values. The number of signs is 15, of which 9 are positive and 6 negative, and these are arranged in six sets. Making the appropriate calculations, we have

$$P\{6 \text{ sets} \mid 9 \text{ positive; } 6 \text{ negative}\} = \tfrac{875}{5005} = 0{\cdot}175.$$

The arrangement of signs will therefore be judged as acceptable. The joint significance of a $P\{\chi^2\} = 0{\cdot}10$ and a $P\{T\} = 0{\cdot}175$ is found, by evaluating the joint distribution, to be 0·066.

Table 6. *Sample values. Observed and expected*

Central values	−2·1 and under	−1·8	−1·5	−1·2	−0·9	−0·6	−0·3	0·0
Observation	12	10	18	26	23	42	43	49
Expectation	9·3	8·6	14·0+	21·0+	28·7	35·9	41·0+	43·0−
Deviation	+2·7	+1·4	+4·0	+5·0	−5·7	+6·1	+2·0	+6·0

Central values	+0·3	+0·6	+0·9	+1·2	+1·5	+1·8	+2·1 and over	Total
Observation	35	28	20	26	20	3	5	360
Expectation	41·0+	35·9	28·7	21·0	14·0	8·6	9·3	360
Deviation	−6·0	−7·9	−8·7	+5·0	+6·0	−5·6	−4·3	0

16. A study of the basic table (Table 1) of the function T will show that $P\{T\}$ is not a very sensitive criterion with which to judge the randomness of a sequence of signs unless the number of groups under consideration is very large. For example, if there are 10 signs, 5 of which are positive and 5 of which are negative, the probability of getting two sets of signs is 0·008. Thus the test would show, and rightly, that the chance of such an arrangement is small, but this fact would undoubtedly be recognized by a skilled computer without the use of a test at all. In the case of 10 signs the probability of three groups or less is 0·040, and this would possibly be judged non-significant. Again, let us consider an extreme case say, 10 signs, 9 of which are positive and 1 negative. The T criterion does not concern itself with the fact that the numbers 9 and 1 are exceptional, it is merely concerned with deciding whether their arrangement is exceptional given the 9 and 1. Table 1 shows that neither possible arrangement would be considered out of the ordinary. It is these points of weakness which show that the criterion T is not of great utility except in combination with χ^2. For, if we consider the 9 positive, 1 negative case, common sense tells us that the χ^2 criterion in such

a case would possibly be significant. Nine positive deviations have to be balanced by a single negative deviation, and this last is therefore likely to be big. This does not influence T; neither will the contribution of T to the joint criterion be of much weight. This is as it should be, for it is difficult to see how one can postulate a smooth alternative for 9 positive, 1 negative, two sets, and not also for 9 positive, 1 negative, three sets. Generally, however, we shall not meet such extreme cases in practice. One way of overcoming this weakness of the test would be to consider the probability of obtaining r_1 positive and r_2 negative signs together with the probability of obtaining T sets of alternate positive and negative signs given r_1 and r_2. This is simple enough when considering just a sequence of alternatives, as I have shown elsewhere, but it is not easy to fit these results to the χ^2 problem, nor, when this is possible, will the choice of a critical region be straightforward. However, the results of sampling experiments will be utilized to throw light on these points and it is hoped to discuss them, with other questions arising, in a further publication.

17. It is possible that there are other criteria, depending on the arrangement of positive and negative signs, which will be more sensitive than the T criterion chosen. For example, it is easy to calculate, given r_1 and r_2, the probability that the largest set is composed of a sequence of r' positive signs, and there are many other possibilities which might be considered. It would appear that any criterion based on sign sequences can be shown to be independent of χ^2 by means of geometrical argument, and it will be necessary therefore to consider the power of these different sign tests when referred to a specified set of alternate hypotheses.

18. The main objection to the two criteria, T and $P\{\chi^2\}.P\{T\}$, that I have proposed in this note is the one which was mentioned earlier; they are only applicable to the case where there is just one restriction on χ^2, i.e. when the totals of expected and observed frequencies have been made to agree. It is possible to work out a slightly different form of the T criterion for each additional restriction which is put on χ^2, and this has been done. It is preferable, however, to delay publication until the results of an extensive sampling experiment are complete in order to verify whether such theoretical assumptions as have been made are reasonable.

REFERENCES

NEYMAN, J. (1937). *Skand. Aktuar. Tidskr.* **20**, 149–99.
NEYMAN, J. & PEARSON, E. S. (1928). *Biometrika*, **20** A, 263–94.

AN EXACT TEST FOR THE EQUALITY OF VARIANCES*

By R. L. PLACKETT, M.A.

Introduction

The problem of testing the equality of variances and covariances in normal distributions is one which has received considerable attention; we have compiled a bibliography of some sixty papers, and shall issue a survey of these in due course; only papers vital to our discussion will be considered here. A precise instance of the type of situation we are considering is as follows: measurements of height, span and tibia length are made on each of 20 Englishmen, 20 Scotsmen, 20 Welshmen and 20 Irishmen; it is required to know if the covariance matrix of the three characteristics is the same for each of the four nationalities. Nothing is known or assumed about the mean values of these characteristics in the four populations considered, nor are we interested in testing any hypothesis concerning the means, although such a hypothesis may be the object of further investigations which assume that the four covariance matrices are the same; this latter assumption is inevitably made in multivariate analysis of variance.

Wilks (1932) has already given the moments of the distribution of his criterion for testing the equality of several covariance matrices (on the hypothesis that the matrices are in fact equal) and Bishop (1939) put this criterion into an approximate workable shape. The test criterion given here differs from that of Wilks and has the advantage when one or two correlated characteristics are being measured (height or height and span, for example) that its distribution is exactly known whatever the number of populations. Nair (1939) did, it is true, give the exact distribution of the Neyman & Pearson (1931) L_1 criterion for one measured characteristic; and the exact distribution for two characteristics of Wilks's generalization of their criterion; but the form in which the distribution was obtained is very involved. It is interesting to notice that from our standpoint the problem of testing the equality of several variances (i.e. the case of one measured characteristic) is, as will appear, brought within the framework of multiple correlation theory. In the general case of more than two characteristics the moments of the distribution of our criterion, like those of Wilks, are available.

Outline of method

In the usual terminology we consider k p-variate normal distributions and are concerned with testing the hypothesis that the corresponding variances and covariances are all equal. The method we employ to test this hypothesis is essentially that which has been in use in analysis of variance since its origination by Fisher; to test the equality of a set of k quantities we test whether $(k-1)$ orthogonal linear functions of the quantities are each zero. To illustrate the application of this principle in the present instance take the particular case $p = 1$, i.e. we wish to test the equality of the variances in k univariate normal distributions. If a typical observation from the lth distribution is t_l $(l = 1, 2, \ldots, k)$, form k mutually orthogonal linear functions of the t_l such that one is

$$u = t_1 + t_2 + \ldots + t_k.$$

* Communication from the National Physical Laboratory.

If the $(k-1)$ covariances of u and each of the other linear functions are all zero then the variances of the k distributions must all be equal; this condition may be expressed by saying that the multiple correlation coefficient of u on the other linear functions is zero. Further, if there are n sample values of u then the size of sample drawn from each distribution must also be n at least, and if no observations are to be discarded the size of each sample must be n exactly. Thus, although it is not a condition of the problem that the sizes of samples drawn from the k distributions must all be equal, it is a condition of our solution.

The extension of the foregoing principle to $p>1$ is straightforward and is considered in detail in the next section; the problem then becomes that of testing the independence of two groups of variates, the first of size p, i.e. p expressions of the form u; and the second of size $p(k-1)$ comprising all the other orthogonal linear functions. This problem has been treated by Wilks (1935, 1943) and the relevant distribution is expressible as an incomplete β-function when $p=1$ and 2 (for all k); an exact distribution is also known when $p=3$ and 4 for $k=2$. Finally, since when $p=1$ the criterion has the form of a multiple correlation coefficient, the power of the test in this instance can be calculated by virtue of the work of Fisher (1928).

Discussion of the test

A sample of n observations is drawn from each of the k p-variate normal distributions of which the lth has the covariance matrix V^l_{ij} $(l=1,2,\ldots,k;\ i,j=1,2,\ldots,p)$. It is required to test the hypothesis that

$$V^l_{ij}=V^m_{ij} \quad (l,m=1,2,\ldots,k). \tag{1}$$

The population means do not enter into the hypothesis and have arbitrary unknown values. Where i, j, l, m appear henceforth they will be understood to range over the values given above unless otherwise stated. The observations may be written in the form of an $n\times kp$ matrix X such that all those on the ith variate in the lth distribution are in column $(i-1)k+l$. The αth observation in this column $(\alpha=1,2,\ldots,n)$ is denoted by $x^l_{i\alpha}$; the order of the elements in a column is assumed to be random. If this is doubted the observations should be randomly rearranged.

We must emphasize here that the sample value of the criterion to be used to test (1) depends on this order, and there is thus, in a sense, a correspondence between $x^l_{i\alpha}$ and $x^m_{j\alpha}$, although these two quantities are, of course, uncorrelated when $l\neq m$. Most tests of a hypothesis specifying nothing about the order in which observations are made or written down are themselves independent of it; ours is not, and different computers with the same data might well come to different conclusions although this does not affect the validity of the test, the significance level being overall what it should be. There is probably some loss of power which can, however, be offset by imbuing α with a certain physical meaning; but we shall not discuss this question here. A criterion for testing normality depending on the order of arrangement of observations has been suggested by R. C. Geary (1935, pp. 316–17).

Let now

$$z^l_{i\alpha}=x^l_{i\alpha}-\left(\sum_\alpha x^l_{i\alpha}\right)\Big/n, \tag{2}$$

and let the corresponding $n\times kp$ matrix be Z. If $G=Z'Z$, where a prime is used to denote the transpose of a matrix, then, apart from a factor n, G is the matrix of sample variances and covariances of all variables. We further define $S(k,p)$ as the sum of all (k^{2p}) signed minors

$$g^{m_1,m_2,\ldots,m_p}_{l_1,l_2,\ldots,l_p},$$

formed by rows $l_1, l_2, \ldots, l_p$ and columns $m_1, m_2, \ldots, m_p$ of G, where

$$(i-1)k < l_i,\ m_i \leqslant ik. \tag{3}$$

$\bar{S}(k,p)$ is similarly defined for the matrix $\bar{G} = G^{-1}$ (we shall use this notation for the inverses of matrices throughout).

We now proceed to prove the following

THEOREM: $$W(k,p) = k^{2p}/S(k,p)\,\bar{S}(k,p)$$

is distributed like Wilks's statistic for testing the hypothesis that two groups of variates of sizes p and $p(k-1)$, known to have been drawn from a (kp)-variate normal distribution, are mutually independent (Wilks, 1935, 1943). If the groups are in fact mutually independent then (1) is true.

Proof. Introduce a $k \times k$ orthogonal matrix B, the elements of whose first column are all equal (to $\pm 1/\sqrt{k}$) but which is otherwise quite arbitrary. Put

$$r = (i-1)k+l, \quad u = (m-1)p+j, \tag{4}$$

and form a $kp \times kp$ matrix A such that

$$a_{ru} = \delta_{ij} b_{lm}, \tag{5}$$

where $\delta_{ij} = 1$ $(i = j)$, otherwise 0. Clearly A is also orthogonal. For example, suppose $k = 4$, $p = 2$. Apart from a factor of $\pm \frac{1}{2}$ multiplying each element, let

$$B = \begin{Vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{Vmatrix}.$$

Then

$$A = \begin{Vmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \end{Vmatrix}.$$

When $p = 1$, $A = B$. Let

$$D = XA, \quad Y = ZA, \quad C = Y'Y = A'GA. \tag{6}$$

Putting $$s = (j-1)k+m, \quad t = (l-1)p+i, \tag{7}$$

and defining $$t' = (l-1)p+i, \quad u' = (m-1)p+j \quad (l, m = 2, 3, \ldots, k), \tag{8}$$

we have $$\mathscr{E}(g_{rs}) = \delta_{lm}(n-1)V^i_{ij}, \tag{9}$$

so that $$\mathscr{E}(c_{iu}) = (n-1)\sum_l b_{lm} V^i_{lj}/\sqrt{k}. \tag{10}$$

Hence when $$\mathscr{E}(c_{iu'}) = 0 \tag{11}$$

equations (1) are satisfied, because for fixed i and j equations (10) can be solved and yield

$$(n-1)V^i_{lj} = \mathscr{E}(c_{ij}). \tag{12}$$

Denote a typical element of the tth column of D by d_t. Then equations (11) are satisfied if and only if d_i and $d_{u'}$ are mutually independent.

A criterion for testing (11), obtained by likelihood-ratio methods, has been given by Wilks (1935, 1943). This is

$$W(k,p) = \frac{|c_{tu}|}{|c_{ij}||c_{ru'}|}, \tag{13}$$

and is sometimes called the vector alienation coefficient. Let $C^{(p)}$ be the pth compound of C (Aitken, 1939, p. 90), i.e. the matrix of all $p \times p$ minors of C; and $\bar{C}^{(p)}$ the pth compound of $\bar{C} = C^{-1}$ (since $\bar{C}^{(p)}$ is the inverse of $C^{(p)}$ our notation is consistent). Then

$$W(k,p) = 1/c_{11}^{(p)}\bar{c}_{11}^{(p)} \tag{14}$$

by an application of Jacobi's theorem on the minors of the adjugate (Aitken, 1939, p. 97). Now by the Binet-Cauchy theorem (Aitken, 1939, p. 93),

$$C^{(p)} = (A')^{(p)} G^{(p)} A^{(p)}, \quad \bar{C}^{(p)} = (A')^{(p)} \bar{G}^{(p)} A^{(p)}. \tag{15}$$

Consider the elements in the first row of $(A')^{(p)}$. The first p rows of A' are of the form

$$\begin{Vmatrix} 11\ldots1 & 00\ldots0 & 00\ldots0 & \ldots & 00\ldots0 \\ 00\ldots0 & 11\ldots1 & 00\ldots0 & \ldots & 00\ldots0 \\ 00\ldots0 & 00\ldots0 & 11\ldots1 & \ldots & 00\ldots0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 00\ldots0 & 00\ldots0 & 00\ldots0 & \ldots & 11\ldots1 \end{Vmatrix}$$

apart from the factor $\pm 1/\sqrt{k}$ multiplying each element. Therefore the only non-zero elements in the first row of $(A')^{(p)}$ are those formed by taking one column from each of the p blocks of k columns into which the first p rows of A' may be divided. All the non-zero elements equal $k^{-\frac{1}{2}p}$. Then from (15)

$$c_{11}^{(p)} = S(k,p)\,k^{-p}, \quad \bar{c}_{11}^{(p)} = \bar{S}(k,p)\,k^{-p}, \tag{16}$$

so finally

$$W(k,p) = k^{2p}/S(k,p)\,\bar{S}(k,p). \tag{17}$$

This completes the proof.

Case of $p = 1$

Here

$$W(k,1) = k^2/S(k,1)\,\bar{S}(k,1),$$

where $S(k,1)$, $\bar{S}(k,1)$ are the sums of all elements of G, G^{-1} respectively. If (1) is true, $\mathbf{W}(k,1)$, the true value of $W(k,1)$, is unity. Define

$$W(k,1) = 1 - R^2 \quad \text{and} \quad \mathbf{W}(k,1) = 1 - \mathbf{R}^2, \tag{18}$$

so that if (1) is true, $\mathbf{R} = 0$. The distribution of $R^2 = 1 - W(k,1)$ when $\mathbf{R} = 0$ is, as Wilks pointed out, well known, being that of the multiple correlation coefficient (of d_1 on $d_2, d_3, \ldots, d_k$); if in the usual notation

$$I_x(a,b) = [B(a,b)]^{-1} \int_0^x x^{a-1}(1-x)^{b-1}\,dx, \tag{19}$$

then the cumulative distribution function of $x = R^2$ is $I_x(k-1, n-k)$, values near 1 being significant; that of $x = W(k,1)$ being $I_x(n-k, k-1)$ with small values significant. Tables in convenient form have been calculated by Thompson (1941); otherwise we can convert to the variance-ratio F by

$$F = (n-k)(1-W)/(k-1)\,W. \tag{20}$$

It is clear that n must exceed k; for p variates, n exceeds pk in order that G may be non-singular.

If the matrix A is defined instead as a $kp \times kp$ orthogonal matrix, the elements of whose first column are all equal (cf. equation (5)), the problem is effectively reduced to the case $p = 1$ whatever the value of p, and we can test exactly the somewhat indefinite hypotheses

$$V_{11}^1 + V_{12}^1 + \ldots + V_{1p}^1 = V_{j1}^m + V_{j2}^m + \ldots + V_{jp}^m. \tag{21}$$

This may be applied in the following manner, for take $k = p = 2$ and obtain

$$V_{11}^1 + V_{12}^1 = V_{21}^1 + V_{22}^1 = V_{11}^2 + V_{12}^2 = V_{21}^2 + V_{22}^2. \tag{22}$$

Thus
$$V_{11}^1 = V_{12}^1 \quad \text{and} \quad V_{11}^2 = V_{22}^2. \tag{23}$$

If it is assumed
$$V_{11}^1 = V_{11}^2, \tag{24}$$

then
$$V_{12}^1 = V_{12}^2, \tag{25}$$

and conversely.

Case of $p = 2$

The distribution of $W(k, 2)$ has been given by Wilks (1935). If $x = \sqrt{[W(k, 2)]}$, the cumulative distribution function of x is

$$I_x(n - 2k, 2k - 2). \tag{26}$$

Small values of x are significant and n must exceed $2k$.

Case of $p \geqslant 3$

For $k = 2$, $p = 3$ and 4, the exact distributions are again known and have been given by Wilks in equations (35) and (37) respectively of his 1935 paper. The expressions are rather complicated and we have not reproduced them here. For other values of k and p the moments of $W(k, p)$ are available; while more recently Wald & Brookner (1941) have obtained the distribution in the form of an infinite series, calculating numerical values for the coefficients in certain instances.

For $p > 1$, (17) becomes rather intractable as a means of calculating $W(k, p)$. Indeed, for $k = 2$ and $p = 4$ it is necessary

(i) to calculate 36 sample variances and covariances,
(ii) find the inverse of an 8×8 matrix,
(iii) calculate 512 4×4 determinants,

and it is clearly better to reintroduce the matrix A in some appropriate numerical form, calculate $Y = ZA$ and $C = Y'Y$, and find $W(2, 4)$ from (13), a process which involves the evaluation of an 8×8 and two 4×4 determinants.

POWER OF THE TEST WHEN $p = 1$

From (17) the true value of R^2 is in general given by

$$1 - \mathbf{R}^2 = k^2 \Big/ \left[\left(\sum_i V_{11}^i \right) \left(\sum_i 1/V_{11}^i \right) \right], \tag{27}$$

and thus the test will have equal power for all values of the variances such that the product of their sum and the sum of their reciprocals is constant. Consequently $1 - W(k, 1)$ is distributed like the multiple-correlation coefficient in samples from a population where the

true value is given by (27). The probability density function of this distribution has been deduced by Fisher (1928) and can be integrated to give a finite series when $(n-k)$ is even.

We find easily when $k = 2$ that in the V_{11}^1, V_{11}^2 quarter-plane the equipotentials are pairs of lines

$$V_{11}^1 = aV_{11}^2 \quad \text{and} \quad aV_{11}^1 = V_{11}^2, \tag{28}$$

where

$$a = (1+\mathbf{R})/(1-\mathbf{R}). \tag{29}$$

For $k > 2$ the equipotential surfaces in k dimensions are cones through the origin situated symmetrically with regard to the co-ordinate primes.

Reverting to $k = 2$ three methods are available for testing the hypothesis that $V_{11}^1 = V_{11}^2$:

(i) Fisher's z or $F = \exp(2z)$

$$= g_{11}/g_{22}. \tag{30}$$

(ii) the L_1 criterion introduced by Pearson & Neyman (1930) and later extended to $k > 2$ (Neyman & Pearson, 1931).

In the instance we are considering, i.e. equal sample sizes from both populations,

$$L_1 = 2(g_{11}g_{22})^{\frac{1}{2}}/(g_{11}+g_{22}) \tag{31}$$

$$= 2F^{\frac{1}{2}}/(1+F). \tag{31a}$$

(iii)
$$W(2,1) = 4[g_{11}g_{22} - (g_{12})^2]/[(g_{11}+g_{22})^2 - (2g_{12})^2].* \tag{32}$$

Thus tests (i) and (ii) are exactly equivalent, as is known, the optimum critical region being that corresponding to equal tails of the F-distribution. Criterion (iii) is that obtained by Morgan (1939) and Pitman (1939), appearing as equation (12) in Morgan's paper, to test that the variances in a normal bivarate population are equal. Morgan has compared the powers of tests (i) and (iii) for $n = 12$, 25 and 100 at a significance level of 0·10 and for these sample sizes it appears that the tests are effectively of equal power.

When n is large and, consequently, the two populations being independent, $(g_{12})^2/g_{11}g_{22}$ is converging in probability to zero,

$$W(2,1) \sim L_1^2. \tag{33}$$

The cumulative distribution functions of criteria (ii) and (iii) are respectively $I_x\left(\frac{n-1}{2}, \frac{1}{2}\right)$ (Nayer, 1936) ($x = L_1^2$) and $I_x\left(\frac{n-2}{2}, \frac{1}{2}\right)$ ($x = W(2,1)$). Generally, $W(k,1)$ for large n converges in probability to the harmonic mean of the sample variances divided by their arithmetic mean; L_1 (for equal sample sizes) is exactly equal to the geometric mean divided by the arithmetic mean.

Example of the use of the test for a case with $k = 4$, $p = 1$

It is not easy to calculate $W(k,1)$ from equation (17) if $k > 3$. Indeed, the main value of (17) lies in showing the form of solution, and in establishing that this is independent of the particular orthogonal transformations used. In the following example, therefore, orthogonal transformations are made at once and the multiple correlation coefficient is calculated from the numerical data. This procedure is far quicker than that involved in calculating $W(4,1)$ from (17).

* See Appendix.

Below are given samples of 10 from each of four univariate normal populations:

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
−20	+24	+ 4	+52	+ 7	+15	+ 8	− 8
− 1	+18	+ 9	−24	+ 5	+24	− 1	+56
−11	+27	−27	0	+18	−12	+ 1	−64
+10	+21	+ 5	+48	+13	−24	− 4	+12
− 4	−48	− 3	+48	− 6	+12	+ 5	−12

which have mean zero and standard deviations respectively 10, 30, 10, 40. Make the following orthogonal transformation:

$$y_1 = x_1+x_2+x_3+x_4, \quad y_2 = x_1+x_2-x_3-x_4,$$
$$y_3 = x_1-x_2+x_3-x_4, \quad y_4 = x_1-x_2-x_3+x_4,$$

and obtain

y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
+60	−52	−92	+ 4	+22	+22	+ 8	−24
+ 2	+32	+14	−52	+84	−26	−76	+38
−11	+43	−65	−11	−57	+69	+95	−57
+84	−22	−54	+32	− 3	−19	+21	+53
− 7	−97	− 7	+95	− 1	+13	− 1	−35

Form the matrix of sums of squares and cross-products, i.e. C. This is

$$\begin{Vmatrix} +18636{\cdot}1 & - 9646{\cdot}9 & -18232{\cdot}9 & + 7325{\cdot}1 \\ - 9646{\cdot}9 & +21784{\cdot}1 & +12018{\cdot}1 & -19015{\cdot}9 \\ -18232{\cdot}9 & +12018{\cdot}1 & +28692{\cdot}1 & - 9445{\cdot}9 \\ + 7325{\cdot}1 & -19015{\cdot}9 & - 9445{\cdot}9 & +22008{\cdot}1 \end{Vmatrix}.$$

The matrix of sample correlation coefficients is therefore

$$\begin{Vmatrix} 1 & -0{\cdot}4788 & -0{\cdot}7885 & +0{\cdot}3617 \\ -0{\cdot}4788 & 1 & +0{\cdot}4807 & -0{\cdot}8685 \\ -0{\cdot}7885 & +0{\cdot}4807 & 1 & -0{\cdot}3759 \\ +0{\cdot}3617 & -0{\cdot}8685 & -0{\cdot}3759 & 1 \end{Vmatrix}.$$

Hence the multiple correlation coefficient of y_1 on y_2, y_3, y_4 is given by $R^2 = 0{\cdot}637$. Calculated by the approximation indicated in the last paragraph of the preceding section* $R^2 = 0{\cdot}656$; the true value obtained from equation (27) with variances in the ratio 1 : 3 : 1 : 4 is 0·727. The upper 10 and 5 % levels of significance, obtained from Thompson's tables with $\nu_1 = n-k = 6$, $\nu_2 = k-1 = 3$, are respectively 0·622 and 0·704. We find $L_1 = 0{\cdot}565$, the 5 and 1 % levels obtained from Nayer's (1936) tables being respectively 0·797 and 0·719, so that this test gives a more significant result than the one based on R^2. The relative merits of L_1 and the test we have provided, which cannot be judged on the results of one example, remain a problem to be investigated.

* I.e. calculated from 1 − (harmonic mean of g_{ii})/(arithmetic mean of g_{ii}).

SUMMARY

An exact test has been put forward for the equality of variances and covariances in any number k of 1- or 2-variate normal populations; the test is also exact for two 3- or 4-variate populations; but is restricted in application to equal sample sizes n from the k populations where n exceeds pk, p being the number of variates. The moments of the criterion are available for k p-variate populations where the statistic used is equivalent to that employed by Wilks (1935) to test the independence of two groups of variates (of sizes p and $p(k-1)$), and has the same distribution. In the univariate case the power of the test is known as a function of one parameter. Comparison with the L_1 criterion has already been made when $p = 1$ and $k = 2$, the tests being practically the same, and an example worked out of the use of the test when $p = 1$.

Our thanks are due to E. C. Fieller for drawing our attention to the papers by Morgan and Pitman and suggesting that the test given there for the equality of two variances might be extended to more than two; also to Prof. E. S. Pearson for pointing out the need of certain explanatory additions.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

AITKEN, A. C. (1939). *Determinants and Matrices.* Oliver and Boyd, Edinburgh.

BISHOP, D. J. (1939). On a comprehensive test of the homogeneity of variances and covariances in multivariate problems. *Biometrika*, **31**, 31.

FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.* A, **121**, 654.

GEARY, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, **27**, 310.

MORGAN, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, **31**, 13.

NAIR, U. S. (1939). The application of the moment function in the study of distribution laws in statistics. *Biometrika*, **30**, 274.

NAYER, P. P. N. (1936). An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits. *Statist. Res. Mem.* **1**, 38.

NEYMAN, J. & PEARSON, E. S. (1931). On the problem of k samples. *Bull. int. Acad. Cracovie*, Série A, p. 460.

PEARSON, E. S. & NEYMAN, J. (1930). On the problem of two samples. *Bull. int. Acad. Cracovie*, Série A, p. 73.

PITMAN, E. J. G. (1939). A note on normal correlation. *Biometrika*, **31**, 9.

THOMPSON, C. M. (1941). Tables of percentage points of the incomplete beta-function. *Biometrika*, **32**, 168.

WALD, A. & BROOKNER, R. J. (1941). On the distribution of Wilks' statistic for testing the independence of several groups of variates. *Ann. Math. Statist.* **12**, 137.

WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471.

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrika*, **3**, 309.

WILKS, S. S. (1943). *Mathematical Statistics.* Princeton University Press.

APPENDIX

As an illustration of the algebraic form of $W(k, 1)$ the Editor has suggested to me that it might be helpful to show the relation of the general formula (17) to the matrix G in this simple case when $k = 2$. Here, using a common notation for a sample mean

$$g_{11} = \sum_{\alpha=1}^{n} (x_{1\alpha}^1 - x_{1\cdot}^1)^2, \quad g_{22} = \sum_{\alpha=1}^{n} (x_{1\alpha}^2 - x_{1\cdot}^2)^2, \quad g_{12} = \sum_{\alpha=1}^{n} (x_{1\alpha}^1 - x_{1\cdot}^1)(x_{1\alpha}^2 - x_{1\cdot}^2),$$

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}, \quad \bar{G} = \begin{bmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{bmatrix} \Big/ \{g_{11}g_{22} - g_{12}^2\},$$

$$S(2, 1) = g_{11} + g_{22} + 2g_{12}, \quad \bar{S}(2, 1) = \frac{g_{11} + g_{22} - 2g_{12}}{g_{11}g_{22} - (g_{12})^2}.$$

Whence, using (17), (32) is at once obtained for $W(2, 1)$. For $k > 2$ the full expression for $\bar{S}(k, 1)$ in terms of the g's is complicated and the matrix notation becomes essential.

THE ESTIMATION FROM INDIVIDUAL RECORDS OF THE RELATIONSHIP BETWEEN DOSE AND QUANTAL RESPONSE

By D. J. FINNEY

Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford

1. Introduction

A type of biometric problem frequently encountered by the statistician is that which requires the estimation and study of a relationship between dose and response. 'Dose' is here a general term indicating the magnitude of a stimulus applied to certain test subjects, and 'response' is a measure of the effect which the stimulus produces on the subjects. When the test subjects are living matter, whether plants, animals or bacteria, pieces of tissue or single cells, the response to a specified dose is unlikely to be constant in repeated trials, and regression methods must be used in the estimation of the relationship.

In some classes of data, the response is 'all-or-nothing' or quantal, and cannot be measured quantitatively. Ordinary regression methods are then no longer applicable; methods based on the transformation of the proportion of subjects showing the response at any dose level to the normal equivalent deviate (Gaddum, 1933), or to the probit (Bliss, 1934*a*, *b*), however, have proved very powerful for simplifying the statistical analysis. In recent years, full accounts of the underlying theory of these transformations, and of their application, have been published by various authors (see, for example, Bliss, 1935*a*, *b*; Finney, 1947, 1948). An additional difficulty sometimes found is that the intensity of the stimulus cannot be *selected* in advance of a test, but can only be *measured* after the test has taken place; only rarely will two or more subjects happen to receive exactly the same dose, and more usually the records consist of a list of doses with, for each, a statement of whether a single subject receiving that dose responded or not.* For example, in some methods for the testing of insecticidal potency, poison bait is offered to individual insects; the dose received by any insect cannot be specified in advance, and must instead be measured as the amount of poison ingested.

Data from experiments of this kind do not give empirical values for the proportion of subjects responding at each dose level, except in the trivial sense that every dose shows either zero or 100 % responding. Nevertheless, as Bliss (1938) has pointed out, the probit method can still be applied to estimation of the dose-response relationship. He has given a numerical example, though without showing full details of the working, but has admitted that assessment of the error of estimation presents some theoretical difficulties (Finney, 1947, § 43). An interesting example of experimental results requiring this type of analysis has recently been brought to the notice of the writer by Mr R. W. Gilliatt. These introduce an additional complication, since the dose is expressed in terms of two measurements, and a probit plane (Finney, 1943) or other bivariate regression function must therefore be estimated. An account of the analysis, with computational details, may help those who have encountered analogous problems in biological or other investigations.

* When response does not involve death or serious alteration of the test subject, one subject may be used many times; the example discussed in this paper is an instance. The form of the data will be the same, though the interpretation may require that tolerance variation between and within subjects be distinguished.

2. THE DATA

Research in human physiology has demonstrated that, under carefully controlled experimental conditions, a transient reflex vaso-constriction in the skin of the digits may follow a single deep breath (Bolton, Carmichael & Stürup, 1936). Gilliatt (1947) has found that the response depends in part on the volume of air taken in by the subject. Plethysmographic measurement of the volume changes in a finger was used to indicate the occurrence of a response, but assessment of the degree of vaso-constriction, in order to relate this to the inspiratory stimulus, was not practicable. Thus the records obtained for each test show only

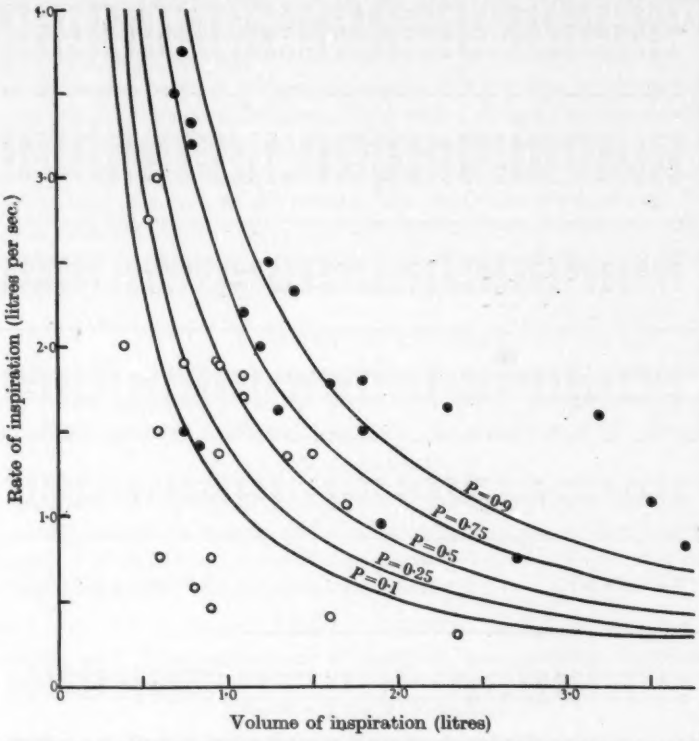


Fig. 1. Contours of dose-response surface for 0·1, 0·25, 0·5, 0·75 and 0·9 frequency of response, estimated from three-parameter equation. ○ no vaso-constriction; ● vaso-constriction.

the volume of air inspired, the average rate of inspiration, and whether or not vaso-constriction was produced. The above brief outline is sufficient for appreciation of the statistical problem, but a full account of the experimental procedure may be found in Gilliatt's paper; the results discussed here are presented in his Fig. 5.

The data, which Mr Gilliatt has kindly made available to the writer, were obtained from thirty-nine tests, in twenty of which vaso-constriction occurred. Tests were made on three different subjects, nine on D.W., eight on V.P.W., and twenty-two on S.J.S.; the results of the tests, with the subjects in this order, are shown in Table 1. In Fig. 1 are shown the thirty-nine combinations of volume in litres (V) and rate of inspiration in litres per second

Table 1. *Experimental data and details of calculations*

Volume in litres (V)	Rate in litres per sec. (R)	x_1	x_2	Response	Y	w	y	wx_1	wx_2	wy	P
3·7	0·825	1·57	0·92	+	6·8	0·18	7·26	0·2826	0·1656	1·3068	0·962
3·5	1·09	1·54	1·04	+	7·3	0·08	7·68	0·1232	0·0832	0·6144	0·989
1·25	2·5	1·10	1·40	+	6·5	0·27	7·02	0·2970	0·3780	1·8954	0·931
0·75	1·5	0·88	1·18	+	3·7	0·34	8·97	0·2992	0·4012	3·0498	0·098
0·8	3·2	0·90	1·51	+	5·8	0·50	6·53	0·4500	0·7550	3·2650	0·789
0·7	3·5	0·85	1·54	+	5·6	0·56	6·42	0·4760	0·8624	3·5952	0·741
0·6	0·75	0·78	0·88	−	1·3	0·00	1·05	0·0000	0·0000	0·0000	0·000
1·1	1·7	1·04	1·23	−	5·1	0·63	3·74	0·6552	0·7749	2·3562	0·530
0·9	0·75	0·95	0·88	−	2·4	0·04	2·06	0·0380	0·0352	0·0824	0·005
0·9	0·45	0·95	0·65	−	1·0	0·00	0·76	0·0000	0·0000	0·0000	0·000
0·8	0·57	0·90	0·76	−	1·4	0·00	1·14	0·0000	0·0000	0·0000	0·000
0·55	2·75	0·74	1·44	−	4·3	0·53	3·53	0·3922	0·7632	1·8709	0·248
0·8	3·0	0·78	1·48	−	4·8	0·63	3·72	0·4914	0·9324	2·3436	0·430
1·4	2·33	1·15	1·37	+	6·6	0·24	7·09	0·2760	0·3288	1·7016	0·950
0·75	3·75	0·88	1·57	+	6·0	0·44	6·66	0·3872	0·6908	2·9304	0·848
2·3	1·64	1·36	1·21	+	7·1	0·11	7·51	0·1496	0·1331	0·8261	0·982
3·2	1·6	1·51	1·20	+	8·0	0·01	8·30	0·0151	0·0120	0·0830	0·999
0·85	1·415	0·93	1·15	+	3·9	0·40	7·87	0·3720	0·4600	3·1480	0·128
1·7	1·06	1·23	1·03	−	5·2	0·63	3·72	0·7749	0·6489	2·3436	0·563
1·8	1·8	1·26	1·26	+	6·7	0·21	7·17	0·2646	0·2646	1·5057	0·958
0·4	2·0	0·60	1·30	−	2·6	0·06	2·23	0·0360	0·0780	0·1338	0·007
0·95	1·36	0·98	1·13	−	4·1	0·47	3·41	0·4606	0·5311	1·6027	0·179
1·35	1·35	1·13	1·13	−	5·1	0·63	3·74	0·7119	0·7119	2·3562	0·533
1·5	1·36	1·18	1·13	−	5·4	0·60	3·62	0·7080	0·6780	2·1720	0·662
1·6	1·78	1·20	1·25	+	6·3	0·34	6·86	0·4080	0·4250	2·3324	0·897
0·6	1·5	0·78	1·18	−	3·0	0·13	2·58	0·1014	0·1534	0·3354	0·025
1·8	1·5	1·26	1·18	+	6·2	0·37	6·79	0·4662	0·4366	2·5123	0·894
0·95	1·9	0·98	1·28	−	5·0	0·64	3·75	0·6272	0·8192	2·4000	0·489
1·9	0·95	1·28	0·98	+	5·2	0·63	6·28	0·8064	0·6174	3·9564	0·577
1·6	0·4	1·20	0·60	−	2·4	0·04	2·06	0·0480	0·0240	0·0824	0·005
2·7	0·75	1·43	0·88	+	5·6	0·56	6·42	0·8008	0·4928	3·5952	0·727
2·35	0·03	1·37	0·48	−	2·8	0·09	2·41	0·1233	0·0432	0·2169	0·015
1·1	1·83	1·04	1·26	−	5·2	0·63	3·72	0·6552	0·7938	2·3436	0·600
1·1	2·2	1·04	1·34	+	5·7	0·53	6·47	0·5512	0·7102	3·4291	0·767
1·2	2·0	1·08	1·30	+	5·8	0·50	6·53	0·5400	0·6500	3·2650	0·776
0·8	3·33	0·90	1·52	+	5·8	0·50	6·53	0·4500	0·7600	3·2650	0·806
0·95	1·9	0·98	1·28	−	5·0	0·64	3·75	0·6272	0·8192	2·4000	0·489
0·75	1·9	0·88	1·28	−	4·3	0·53	3·53	0·4664	0·6784	1·8709	0·243
1·3	1·625	1·11	1·21	+	5·4	0·60	6·34	0·6660	0·7260	3·8040	0·665
					Totals	14·29		14·9980	17·8375	74·9914	
					Means			1·0495	1·2483	5·2478	

Swx_1^2	Swx_1x_2	Swx_2^2	Swx_1y	Swx_2y	Swy^2
16·235606	18·338532	22·783383	79·738989	94·125032	433·566326

(R), together with indications of whether or not the subject responded under these conditions. Inspection of Fig. 1 shows that, in general, when both V and R were small no response occurred, when either was large (unless the other was very small) the response occurred, and in an intermediate region the proportion of responses increased as either V or R increased. There was no sharply defined threshold separating combinations of V and R giving the response from those giving no response; instead, there appeared to be a probability of response ranging from practical certainty under some conditions to zero under others.

As an aid to fuller understanding of the influence of breathing on vaso-constriction, examination of the relationship between V, R, and the probability of response seemed desirable. Since so few observations were available for each subject, the data were unlikely to be sufficient to show differences between subjects; this point is discussed later, but in the main analysis the distinction between subjects is ignored. For any form of response assessment, the testing of one subject many times must introduce a danger that the result of one test will be affected not only by its own stimulus but by preceding stimuli and by the effects they produced. In this investigation, each subject was given a number of preliminary tests until he appeared to have settled into the routine. The observations recorded in Table 1 were obtained after these preliminary trials; they are tabulated in the order of testing, and show no indication of effects of previous history, but clearly such effects would have to be very pronounced if they were to be detectable on this amount of data.

3. METHOD OF ANALYSIS

Preliminary examination of the data suggested that the occurrence of a response was largely determined by the magnitude of VR, the product of volume and rate, curves on which the probability of response has a constant value being approximately hyperbolae of the form

$$VR = \text{constant}. \tag{1}$$

A little consideration shows that an equation of this type is more reasonable than an equation linear in V and R, though the data are almost certainly inadequate for discriminating between many alternative types of relationship that might be postulated. A system of curves similar to, but rather more general than, equation (1), namely,

$$V^{\beta_1} R^{\beta_2} = \text{constant}, \tag{2}$$

was selected for trial; this equation may alternatively be regarded as representing a series of parallel linear relationships

$$\beta_1 \log V + \beta_2 \log R = \text{constant} \tag{2a}$$

between the logarithms of volume and rate for a fixed probability of response.

A specified combination of V and R will not necessarily always give the same result (response or no response) with a subject, for, even though the subject is unaltered, minor uncontrolled variations in his environment may affect his susceptibility to the applied stimulus. For a particular value of V, the threshold value of R (the value which under the conditions prevailing at any instant would be just sufficient to produce a response) will have a frequency distribution; similarly, for a particular R, there will be a frequency distribution of threshold values of V. If these distributions may be taken as normal in $\log V$ and $\log R$, and, for simplicity, they are supposed to be such that the mean of either logarithm is linearly related to the selected value of the other, then the probability of response will be determined by an expression of the form

$$\beta_1 \log V + \beta_2 \log R,$$

and the threshold values of this quantity will be normally distributed. If x_1 and x_2 are written for $\log(10V)$ and $\log(10R)$ respectively (the factor of 10 is introduced in order to make x_1 and x_2 always positive), this statement enables the probability of response, P, to be expressed as

$$P = \int_{-\infty}^{\alpha+\beta_1 x_1+\beta_2 x_2} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du, \tag{3}$$

where α, β_1 and β_2 are parameters to be estimated from the data. The estimation may be regarded as the fitting of a probit regression plane, for Y, the probit of P being given by

$$Y = 5+\alpha+\beta_1 x_1+\beta_2 x_2. \tag{4}$$

Substitution of the value of Y corresponding to a specified probability gives the required linear relationship, equation (2a), between x_1 and x_2 for that probability, from which the estimated curves of constant probability, equation (2), may easily be derived.

The procedure for fitting a probit plane has been described elsewhere (Finney, 1943, 1947, § 31), and its chief features need no alteration for application to individual records. Providing that a first approximation to the equation can be guessed, repeated cycles of computation will give values for the parameters which approach more and more closely to the maximum likelihood estimates. Care in the choice of the first approximation will reduce the number of cycles needed; a poor choice will delay the convergence, though it will not affect the ultimate result. Since only a single observation is available for each combination of x_1 and x_2, every working probit is either a maximum or minimum value, according to whether or not the response occurs. When there is only one dose factor, in the fitting of a probit regression line to records of individuals, grouping of doses and treatment of the observations in a group as if they related to an average dose may reduce the labour of the early computing cycles, but, since it will tend to give an underestimate of the regression coefficient, the final cycle may need to use the detailed records. Bliss (1938) has given an example illustrating grouping of this kind. Grouping is less easily applied, however, when two or more dose factors have to be used, and, for the data under discussion, the individual records were used throughout except in the formation of the first approximation.

In the standard form of probit analysis, with moderately large numbers of observations at each level of dose, a χ^2 is usually computed for testing the significance of discrepancies between the data and the fitted equation; this χ^2 is numerically the same as would be obtained by calculation from expected and observed numbers of responses and non-responses for each dose. If there are few observations in any dose group, the expected number of responses or of non-responses (or of both) is likely to be small, and, as is well known, χ^2 may then fail to follow the sampling distribution tabulated for that statistic. Data of the type under discussion here are extreme examples of this situation, the number of observations for each dose being reduced to unity, so that any disturbance of the χ^2 distribution is likely to be encountered in its most acute form. No complete theoretical investigation of this matter has yet been made, but the practical implications are discussed more fully in § 5.

On the assumption that the estimate of equation (4) is an adequate representation of the data, lines of constant response probability may be obtained for any specified probability; these may be plotted according to equation (2) on a V, R scale. Standard statistical processes also enable fiducial limits to be assigned to the position of any of these curves. The difficulty of dealing with the estimation of error for individual records, and the inadequacy of the data for any sensitive test of whether equation (2) is a satisfactory representation of the

system of curves, throw doubts on the exact interpretation of these fiducial limits. Nevertheless, they give some idea of the confidence that can be attached to the estimated curves, at least for moderate values of V and R; for extremes of either measurement, far more extensive data would be needed before much faith could be placed in the fitted equation.

4. COMPUTATIONS FOR ESTIMATING THE THREE-PARAMETER EQUATION

In this and the two succeeding sections, the computations for Gilliatt's data will be described in detail. The first five columns of Table 1 show the thirty-nine pairs of values of V and R which occurred in the experiments, followed by the corresponding values of x_1 and x_2, together with a statement of whether or not the subject responded. Before the probit computations could be initiated, a first approximation to equation (4) was needed; this was obtained with the aid of the suggestion, from the plotting of the data shown in Fig. 1, that the constant probability curves were approximately the hyperbolae of equation (1), or alternatively

$$x_1 + x_2 = \text{constant}.$$

As Bliss (1938) has pointed out, there is no objection to the use of overlapping groups in the formation of the first approximation. The data were therefore grouped according to the value of $(x_1 + x_2)$, as shown below, and the proportion of responses in each group was obtained from Table 1:

x_1+x_2	Responses	Proportion (p)	Probit of p	First approximation
1·5–1·9	0/7	0·00	—	3·3
1·6–2·0	0/7	0·00	—	3·6
1·7–2·1	2/7	0·29	4·4	3·9
1·8–2·2	2/9	0·22	4·2	4·2
1·9–2·3	3/14	0·21	4·2	4·5
2·0–2·4	8/19	0·42	4·8	4·8
2·1–2·5	13/24	0·54	5·1	5·1
2·2–2·6	17/25	0·68	5·5	5·4
2·3–2·7	16/17	0·94	6·6	5·7
2·4–2·8	12/12	1·00	—	6·0

Each proportion was regarded as an estimate for the median value of $(x_1 + x_2)$ in the group, i.e. 1·7, 1·8, 1·9, ..., and its probit was read from one of the standard tables (Finney, 1947, Table I; Fisher & Yates, 1947, Table IX). As may be seen above, these probits were fairly well fitted by the guessed equation

$$Y = -1{\cdot}8 + 3(x_1 + x_2), \tag{5}$$

which was therefore used as a first approximation to equation (4).

A first set of expected probits was calculated from equation (5), and inserted as Y in an earlier version of Table 1. A cycle of routine probit calculations, just as described in the next two paragraphs, then led to an improved approximation to the required estimate, on which a second cycle of improvement was based. The figures shown in Table 1 relate to the fourth of these cycles, based upon the approximation

$$Y = -9{\cdot}127 + 6{\cdot}666x_1 + 5{\cdot}906x_2 \tag{6}$$

from the third cycle. Equation (6) is very different from equation (5), suggesting that more care might have been given to the selection of a first approximation; that the grouping

adopted would lead to underestimation of the regression coefficients was expected, but insufficient allowance for this was made. Of course the 'improvement' in the approximations refers to their approach to the solution of the maximum likelihood equations, and is not necessarily always an approach to the true relationship.

The column of expected probits, Y, in Table 1 was calculated by substitution of pairs of values x_1, x_2 in equation (6); one decimal place here is quite sufficient. The weighting coefficient, w, for each observation was then read from tables (Finney, 1947, Table II; Fisher & Yates, 1947, Table XI) and entered in its column. The working probit, y, takes a maximum value for every observation giving a response and a minimum value for every observation giving no response, since these give empirical rates of 100 % and zero respectively; values of y were read directly from Finney's table (1947, Table III; or, less simply for the minimum values, from Fisher & Yates, 1947, Table XI). The numbers of decimal places shown for the entries in Table 1 are sufficient for data of this type; indeed possibly one decimal for w and for y would be enough. Columns wx_1, wx_2, and wy were then filled, and the weighted sums of squares and products of deviations, required for the calculation of the regression of y on x_1 and x_2, were completed at the bottom of the table.

The equations giving the estimates of the regression coefficients, b_1 and b_2, are

$$0{\cdot}494528b_1 - 0{\cdot}382729b_2 = 1{\cdot}032130,$$
$$-0{\cdot}382729b_1 + 0{\cdot}517714b_2 = 0{\cdot}516978.$$

Later calculations use the variances and covariance of b_1 and b_2; the equations were therefore solved by first obtaining the matrix inverse to that formed by the coefficients of b_1 and b_2 (Finney, 1943, 1947, § 31; Fisher, 1946, § 29). This matrix is

$$V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} = \begin{pmatrix} 4{\cdot}726144 & 3{\cdot}493883 \\ 3{\cdot}493883 & 4{\cdot}514482 \end{pmatrix}; \tag{7}$$

the accuracy of the data is insufficient to need the number of decimal figures shown here, but their retention assists the checking and maintains the internal consistency of the analysis. Now

$$b_1 = 1{\cdot}032130v_{11} + 0{\cdot}516978v_{12}$$
$$= 6{\cdot}68426,$$

and similarly $b_2 = 5{\cdot}94003$.

The estimate of equation (4) is then

$$Y = \bar{y} + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2)$$

or

$$Y = -9{\cdot}182 + 6{\cdot}6843x_1 + 5{\cdot}9400x_2, \tag{8}$$

a result which differs little from equation (6) and may be regarded as a sufficiently close approximation to the maximum likelihood estimate. Since

$$b_2/b_1 = 0{\cdot}889, \tag{9}$$

equation (8) may be transformed to give

$$VR^{0{\cdot}889} = \text{constant} \tag{10}$$

as the relationship estimated to exist between V and R for a specified probability; the value of the constant can be obtained by substitution of the probit of the probability in equation (8), a process which gives 1·10, 1·36, 1·71, 2·16 and 2·66 for probabilities of 10, 25, 50, 75 and 90 % respectively. Typical contours have been drawn in Fig. 1 so as to indicate the form of the relationship.

5. GOODNESS OF FIT

When probit analysis is applied to data containing many observations in each dose group, the weighted sum of squares of deviations between the empirical probits and the predictions from the fitted equations is a χ^2, with degrees of freedom equal to the number of dose groups reduced by the number of fitted parameters. If S_{uv} is written for the weighted sum of products of deviations of variates u and v, application of this method here would give

$$\begin{aligned}\chi^2_{[36]} &= S_{yy} - b_1 S_{x_1 y} - b_2 S_{x_2 y} \\ &= 40{\cdot}045 - 6{\cdot}6843 \times 1{\cdot}0321 - 5{\cdot}9400 \times 0{\cdot}5170 \\ &= 30{\cdot}08. \end{aligned} \tag{11}$$

When the dose groups are small, however, the χ^2 so calculated cannot be trusted as an indicator of the significance of deviations from the fitted equation, and it is presumably most unreliable when each group is reduced to a single observation. Apart from slight discrepancies caused by imperfect approximation to the maximum likelihood solution, the χ^2 in equation (11) is algebraically identical with that which would be derived, by the usual form of calculations, from comparison of observed numbers responding and not responding in each group with expectations computed from the fitted equation. As is well known from the study of contingency tables, when the expectations in some classes are small the sampling distribution of such a χ^2 may be very different from that shown in the standard tables (Finney, 1947, Table VI; Fisher & Yates, 1947, Table IV); with data from individual records, no class can have an expectation greater than unity, and for many the expectation will be very much less, so that the discrepancy from the tabulated χ^2 distribution is likely to be serious.

The general effect of small expectations on the random sampling distribution of χ^2 appears to be that the mean value remains about equal to the number of degrees of freedom, but that the variance in repeated sampling is increased. Consequently, samples from a population according with the null hypothesis are likely to show an excess of very high and very low values, as judged by the tables of χ^2. Thus there is little danger that significant evidence of deviations from expectation will be overlooked in an uncritical application of the test, though apparently significant values of χ^2 need to be examined with care before they are regarded as evidence sufficient to justify rejection of the null hypothesis. Low values, as in Gilliatt's data 30 with 36 degrees of freedom, need cause little alarm, for they clearly indicate no serious deviation from expectation. High values may in the first instance be compared with the standard tables of the χ^2 distribution; if they fall beyond the significance level, a closer examination should be made before judging the null hypothesis to be untenable, for the apparent significance may be due to large contributions from one or two aberrant points. Gilliatt's data provide an illustration of this. The expected probits for each pair of values of x_1 and x_2 in Table 1 have been calculated from equation (8), and the probabilities, P $(= 1 - Q)$, corresponding to these have been entered in the last column of the table; P is then the expectation of the number of responses for each dose. The χ^2 obtained from the observed and expected numbers in seventy-eight classes is easily seen to be the sum of Q/P for all doses giving a response, plus P/Q for all giving no response. Inspection of the column for P shows small contributions to χ^2 everywhere, except for two instances of responses with probabilities of only 0·098 and 0·128, contributing 9 and 7 respectively; clearly the occurrence of these two responses as the most extreme events in thirty-nine trials need not be regarded

as serious evidence against the null hypothesis. The result of calculating χ^2 by this more laborious process is a total of 30·3, which agrees closely with that already given in equation (11).

One method of modifying a χ^2 test so as to remove its extreme sensitivity to deviations from small expectations is to combine expected and observed frequencies over several adjacent groups, so as to obtain groups with larger expectations; the number of degrees of freedom is then taken as the number of remaining groups less the number of fitted parameters. Of course the groups must be chosen objectively, and without regard to the agreement between the frequencies. The statistic still will not follow the χ^2 distribution exactly, but the approximation should be fairly satisfactory under the usual restriction that the groups be so chosen that none of the expected frequencies is small. This procedure often has to be adopted in probit analysis because of small expectations at very low or very high doses (Finney, 1947, § 18). With individual records, however, only very extensive grouping will give expectations sufficiently large for the χ^2 test to be trusted; the reduction of a large χ^2 to a value below the significance level might then appear indicative of an insensitive test rather than of absence of serious discrepancies.*

Probably no completely satisfactory solution of the difficulty is to be expected. Individual records usually arise from experimental work in which the obtaining of large numbers of observations presents considerable difficulty. Often the whole series will consist of less than fifty observations, and, unless previous information enables the range of doses to be chosen satisfactorily, many of the observations will be made at doses for which response is either almost certain or almost impossible. Even if the individual dose-tolerances could be measured directly, a test of normality of their distribution (which is what the χ^2 test attempts to provide) could not be very sensitive when based on only fifty measurements; if, instead, only quantal data are available, indicating merely whether a dose is below or above the tolerance value, a sensitive normality test is still less likely to exist (Finney, 1947, § 43).

Gilliatt's data, a series of only thirty-nine observations, provide an extreme instance of the difficulty of formulating a sensitive test of goodness of fit. Nevertheless, an attempt has been made to examine the discrepancies between the observations and the null hypothesis expressed by equation (4). In Table 2 are compared the observed and expected frequencies when the data are grouped according to the value of $VR^{0\cdot889}$. This is equivalent to a grouping based on the value of Y, the expected probit in equation (8), and, as this quantity had been evaluated for each observation in order to give P, it was used in the construction of Table 2. Since three parameters have been estimated from the data, four groups is the least number for giving a χ^2 test. The limits of the groups were chosen so as to give similar numbers of observations in each. Inspection of Table 2 shows that the groups are still too small for a χ^2 test to be trusted, thus suggesting that the data are inadequate for any useful test of goodness of fit to be made. The only anomaly in Table 2 is the occurrence of two responses where the expectation is 0·3, and this is clearly insufficient to cause much worry.

The inadequacy of the data for detecting any differences in sensitivity between the three subjects may be seen from Table 3. The first nine entries in Table 1 relate to D.W., and sum-

* In his discussion of the analysis of individual records, Bliss (1938) suggests adjustment of the χ^2 test, not by altering the calculation of the statistic but by reducing the number of degrees of freedom allotted to it; he gives an empirical rule for the reduction, based upon the expectations in terminal dose groups. This method, however, not only lacks any theoretical basis, but seems liable to have an effect opposite to that which is needed; it will attribute significance to high values of χ^2 even more readily than will the unadjusted test.

mation of the values of P gives the expected number of responses for this subject; similarly the next eight and the last twenty-two entries give the numbers for V.P.W. and S.J.S. respectively. Inspection of Table 1 shows that the tests on each subject were fairly widely distributed over the range of values of x_1 and x_2. Table 3 shows excellent agreement between totals of observed and expected responses for each subject, thus suggesting that any individual differences that exist are small by comparison with the variation in sensitivity of the same subject in different tests.

Table 2. *Comparison of observed and expected frequencies of response*

Range of Y	Frequencies of results				
	Observed		Total	Expected	
	−	+		−	+
−4	8	2	10	9·72	0·28
4–5	6	0	6	3·92	2·08
5–6	5	8	13	4·26	8·74
6–	0	10	10	0·59	9·41
Total	19	20	39	18·49	20·51

Table 3. *Comparison of subjects*

Subject	Frequencies of results				
	Observed		Total	Expected	
	−	+		−	+
D.W.	3	6	9	4·0	5·0
V.P.W.	4	4	8	3·5	4·5
S.J.S.	12	10	22	11·0	11·0
Total	19	20	39	18·5	20·5

6. LIMITS OF ERROR

The variances of b_1 and b_2 and the covariance between them are respectively v_{11}, v_{22} and v_{12} as defined in equation (7). Hence the variance of Y, the expected probit corresponding to any pair of values x_1, x_2, is

$$V(Y) = \frac{1}{Sw} + v_{11}(x_1 - \bar{x}_1)^2 + 2v_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + v_{22}(x_2 - \bar{x}_2)^2, \qquad (12)$$

where Sw is the sum of the w column in Table 1. All these variances are derived from binomial probability distributions. In the usual form of probit analysis, with a batch of subjects at each dose, the precision of the estimated relationship between dose and response is discussed as though the variation were normal, an assumption which is justifiable on account of the large numbers of individuals involved. Here, with only thirty-nine observations in all, the

assumption is less safe, but may be adopted for lack of any more trustworthy method of dealing with the data. It is unlikely to be seriously misleading, except possibly for extreme levels of the response probability, P.

Equation (12) may now be used in the assignment of fiducial limits to any one of the curves of equal probability given by equation (10). For suppose that t is the normal deviate corresponding to the significance level to be used in defining the fiducial limits, and that Y_0 is the probit of a probability P_0. Then for any values of x_1, x_2 for which

$$(Y-Y_0)^2 > t^2 V(Y),$$

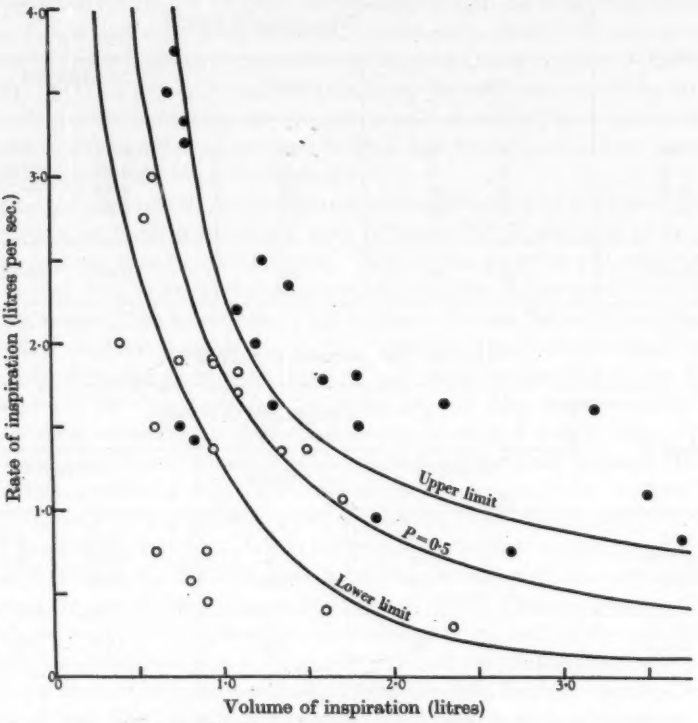


Fig. 2. Fiducial limits (5 % probability) to 0·5 frequency contour of Fig. 1.
○ no vaso-constriction; ● vaso-constriction.

where Y is determined from equation (8), the expected probit differs significantly from Y_0, and for values of x_1, x_2 which reverse the inequality the difference is not significant. Therefore the equation

$$(Y-Y_0)^2 = t^2 V(Y) \qquad (13)$$

gives the limiting values of (x_1, x_2) for which the null hypothesis that the true expected probit is Y_0 is not untenable in the light of the data; in other words, equation (13) defines curves in the (x_1, x_2) plane which are fiducial limits to the estimated locus of points having a constant response probability P_0. These curves are clearly hyperbolae. In Figs. 2 and 3, the 5 % fiducial limit curves ($t = 1{\cdot}960$) for $P_0 = 0{\cdot}5$ and $P_0 = 0{\cdot}9$ respectively have been plotted in

the (V, R) plane; details of the calculation need not be given here, but Fig. 2, for example, is derived from the equation

$$(14{\cdot}182 - 6{\cdot}6843x_1 - 5{\cdot}9400x_2)^2 = 3{\cdot}841\left[\frac{1}{14{\cdot}29} + 4{\cdot}7261(x_1 - 1{\cdot}0495)^2 + 6{\cdot}9878(x_1 - 1{\cdot}0495)(x_2 - 1{\cdot}2483) + 4{\cdot}5145(x_2 - 1{\cdot}2483)^2\right].$$

The pairs of curves are like hyperbolae in form. That for $P_0 = 0{\cdot}5$ defines a band on either side of the estimated relationship which is quite narrow for moderate values of V and R

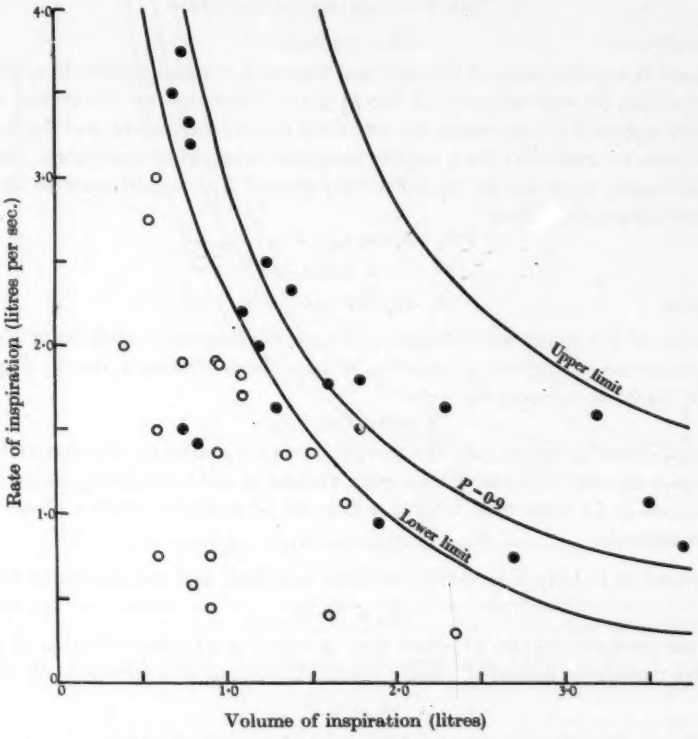


Fig. 3. Fiducial limits (5 % probability) to 0·9 frequency contour of Fig. 1.
○ no vaso-constriction; ● vaso-constriction.

though naturally it widens considerably at the extremes. That for $P_0 = 0{\cdot}9$, as might be expected from general consideration of the problem, allows much greater uncertainty on the side of high values of V and R; similarly, for $P_0 = 0{\cdot}1$, that band would be relatively wider on the side of low values of V or R.

The curves shown in Figs. 1, 2 and 3 may be regarded as plane sections, for selected values of Y, of a three-dimensional diagram relating Y to V and R. In terms of x_1, x_2 instead of V, R, this diagram is the three-dimensional analogue of the familiar diagram showing a regression line with hyperbolic curves indicating limits of error on either side; the line

generalizes to a plane, and the limits are now defined by two sheets of a hyperboloid, one above and one below the plane.

The theoretical basis of the curves illustrated in Figs. 2 and 3 is perhaps insecure, but undoubtedly they give a useful indication of the dependence of the probability of response on V and R and of the reliability of the estimation of this relationship. Much as an experimenter might wish for a more precise assessment of the effects of V and R, experience suggests that results such as those obtained here are as good as can be expected from a total of thirty-nine quantal observations.

7. The two-parameter equation

In § 3, the equation

$$VR = \text{constant} \tag{1}$$

was suggested as an expression of the curves of constant response probability, but the more complex equation (2) was adopted for use in §§ 4–6. There are no theoretical reasons for believing that equation (1) represents the true form of the relationship, and the more general form was chosen in order that the complete calculations might be illustrated. The values of b_1 and b_2 obtained, however, do not differ very greatly by comparison with the standard error of their difference; in fact

$$\begin{aligned} V(b_1-b_2) &= v_{11}-2v_{12}+v_{22} \\ &= 2{\cdot}253, \end{aligned}$$

and therefore

$$b_1-b_2 = 0{\cdot}744 \pm 1{\cdot}501.$$

In the absence of any significant difference between the regression coefficients, the common scientific procedure of preferring the simpler hypothesis (Occam's Razor) suggests that equation (4) might be replaced by

$$Y = \alpha+\beta(x_1+x_2). \tag{14}$$

For the estimation of equation (14), the computations are similar to, but shorter than, those of § 4, since (x_1+x_2) may be replaced by a single variate, x, and a simple regression calculated; the calculations in § 4 were used to give a first set of expected probits, from which was derived the estimate

$$Y = -9{\cdot}475+6{\cdot}4067(x_1+x_2). \tag{15}$$

Only two parameters have been estimated from the data, and calculation as for equation (11) gives

$$\chi^2_{[37]} = 28{\cdot}76.$$

The difference between the two χ^2 values may be taken as a further criterion of whether or not the extra parameter is needed, closely related to the test of significance of (b_1-b_2);

$$\chi^2_{[1]} = 1{\cdot}32$$

is not significant, though again the validity of the χ^2 test is in doubt.

Substitution of the probit of a specified probability in equation (15) gives the value of the constant in equation (1). For the 50 % response probability, for example, the constant is 1·82; over the range of values tested, the curves

$$VR^{0{\cdot}889} = 1{\cdot}71 \quad \text{and} \quad VR = 1{\cdot}82$$

differ only slightly. Similarly, fiducial limits to (x_1+x_2) may be calculated, for any Y_0, as upper and lower values of the product VR. No special interest attaches to these calculations; the novelties due to the individual records are exactly as for the three-parameter equation discussed in earlier sections, and otherwise the method is entirely that of ordinary probit analysis (Finney, 1947, Chapter 4). For comparison with the three-parameter equation, diagrams similar to Figs. 2 and 3 may be prepared; both the constant probability curves and

the fiducial limits are then true hyperbolae. Fig. 4 shows the results for a 50 % response probability, and is to be compared with Fig. 2. The constant probability curve in Fig. 4 differs little from that in Fig. 2, though naturally the difference increases for large values of V or R where the curves are less well determined. For moderate values of V and R, the fiducial

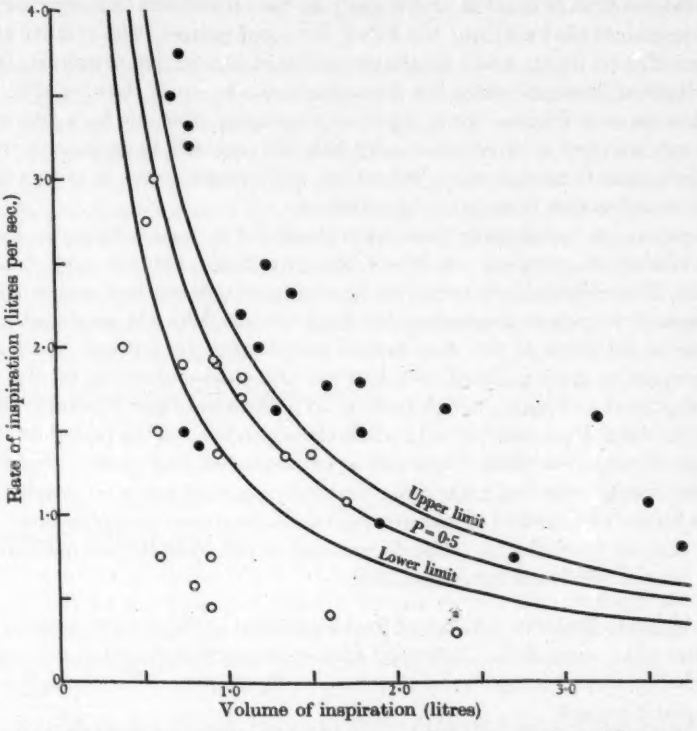


Fig. 4. Contour of dose-response surface for 0·5 frequency of response, estimated from two-parameter equation, and its 5 % fiducial limits (compare Fig. 2). ○ no vaso-constriction; ● vaso-constriction.

limit curves are practically the same as the corresponding curves in Fig. 2, but for more extreme values they lie much closer to the curve of constant probability; since the data show no significant difference between b_1 and b_2, it is to be expected that a more precisely estimated relationship between stimulus and response will be obtained if an assumption that $\beta_1 = \beta_2$ is made, so that the information on the two regression coefficients can be combined, and this shows itself by narrowing the zone of error for the constant probability curve.

8. SUMMARY

The method of probit analysis has been developed to assist the study of the relationship between the magnitude of a stimulus and the proportion of tests in which a particular quantal response to that stimulus appears. In some research problems, the stimulus cannot be controlled sufficiently to make possible the administration of a specified magnitude, though the stimulus actually received by any one subject can later be measured. It will then seldom happen that two subjects receive exactly the same 'dose', and the data for

statistical analysis will generally consist of a series of doses with, for each, a statement of whether or not a single subject showed the characteristic response.

Even for data of this type, the probit transformation can aid the estimation of the relationship between dose and the probability of response. The calculations leading to the estimate are more tedious than is usual in probit analysis, because of slow convergence from a provisional equation to the final form, but follow the usual pattern. The validity of the χ^2 test of goodness of fit (in reality a test for the normality of distribution of individual tolerances) must be doubted, however, since the disturbance due to small class numbers will be encountered in its most extreme form. Extensive grouping of results for adjacent doses will provide a test less open to objection, though this will generally be insensitive to all but the grossest deviations from normality; indeed, no valid sensitive test is to be expected with individual records unless these are very numerous.

In this paper, the calculations have been illustrated on data relating to a reflex vasoconstriction which sometimes occurs in the skin of the digits of human subjects after a single deep breath. The relationship between the occurrence of this response and two dose factors, the volume and the rate of inspiration, has been estimated for the combined records from three subjects; inclusion of two dose factors complicates the analysis, since a bi-variate regression equation must be fitted, but does not affect the underlying theory. The χ^2 test has been discussed at length, though there is no indication of non-normality or of heterogeneity of the data. The reliability with which the dependence of the probability of response on the dose factors is estimated has also been examined, and curves bounding fiducial regions, within which the true probability contours may confidently be asserted to lie, have been determined. This method of representing the limits of error is applicable to other forms of probit analysis involving two dose factors and is not restricted to individual records, though it has not previously been described.

I am indebted to Mr R. W. Gilliatt, of the Department of Physiology, both for permission to make use of his data in an illustration of the statistical methods of my paper and for assistance in describing his experimental procedure. My thanks are due also to Miss M. Callow, who prepared Figs. 1–4.

REFERENCES

Bliss, C. I. (1934*a*). The method of probits. *Science*, **79**, 38–9.
Bliss, C. I. (1934*b*). The method of probits—a correction. *Science*, **79**, 409–10.
Bliss, C. I. (1935*a*). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22**, 134–67.
Bliss, C. I. (1935*b*). The comparison of dosage-mortality data. *Ann. Appl. Biol.* **22**, 307–33.
Bliss, C. I. (1938). The determination of dosage-mortality curves from small numbers. *Quart. J. Pharm.* **11**, 192–216.
Bolton, B., Carmichael, E. A. & Stürup, G. (1936). Vaso-constriction following deep inspiration. *J. Physiol.* **86**, 83–94.
Finney, D. J. (1943). The statistical treatment of toxicological data relating to more than one dosage factor. *Ann. Appl. Biol.* **30**, 71–9.
Finney, D. J. (1947). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve.* Cambridge: University Press.
Finney, D. J. (1948). The principles of biological assay. *J.R. statist. Soc. Suppl.* **9**, 46–91.
Fisher, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Edinburgh: Oliver and Boyd.
Fisher, R. A. & Yates, F. (1947). *Statistical Tables for Biological, Agricultural and Medical Research*, 3rd ed. Edinburgh: Oliver and Boyd.
Gaddum, J. H. (1933). Reports on biological standards. III. Methods of biological assay depending on a quantal response. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 183.
Gilliatt, R. W. (1947). Vaso-constriction in the finger following deep inspiration. *J. Physiol.* (in the Press).

A POWER FUNCTION FOR TESTS OF RANDOMNESS IN A SEQUENCE OF ALTERNATIVES

By F. N. DAVID

1. During recent years attention has been focused on what might be called the 'group' test for randomness in a sequence of alternatives. Thus, if E denote the happening of an event, and $\bar{E}$ its negation, the number of alternations of E and $\bar{E}$ in a sequence supposedly random has been chosen as a test criterion. This test has been put to different uses by W. L. Stevens (1939), A. Wald & J. Wolfowitz (1940) and F. N. David (1947). It seems worth while therefore to enquire what is the power of this test against a set of specifically defined alternate hypotheses. The hypothesis to be tested will be that there is randomness within the sequence, with the alternate hypothesis that if there is no randomness then there is dependence of the type found in a simple Markoff chain. The same procedure will hold good for dependence of the types found in more complex chains although in these cases the enumeration is a little troublesome.

2. If there is a sequence of dependent events

$$E_1, E_2, E_3, \ldots, E_n,$$

then it is an elementary proposition of the probability calculus that

$$P\{E_1 E_2 E_3 \ldots E_{n-1} E_n\} = P\{E_1\} P\{E_2 \mid E_1\} P\{E_3 \mid E_1 E_2\} \ldots P\{E_n \mid E_1 E_2 \ldots E_{n-1}\}.$$

If the events are independent, then

$$P\{E_1 E_2 E_3 \ldots E_{n-1} E_n\} = P\{E_1\} P\{E_2\} P\{E_3\} \ldots P\{E_n\}.$$

This relation will be the basis of H_0, the hypothesis to be tested. If there is dependence as in a simple Markoff chain, then mathematically each event will be dependent on the event immediately preceding it, but will be independent of any of the other events. In this case we shall have

$$P\{E_1 E_2 E_3 \ldots E_{n-1} E_n\} = P\{E_1\} P\{E_2 \mid E_1\} P\{E_3 \mid E_2\} \ldots P\{E_n \mid E_{n-1}\}.$$

This relation will be the basis of H_1, the hypothesis alternate to H_0.

3. For the hypothesis, H_0, let the probability that an event E will occur in a single trial be p, and let the probability of $\bar{E}$ (the negation of E), be q, where $p+q=1$. The probability of obtaining any given sequence of r_1 E's and r_2 $\bar{E}$'s will be

$$p^{r_1} q^{r_2}.$$

The number of ways in which r_1 E's and r_2 $\bar{E}$'s may be arranged to form $2t$ and $2t+1$ sets of E's and $\bar{E}$'s alternately is

$$f_{2t} = \frac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!} \quad \text{and} \quad f_{2t+1} = f_{2t} \times \frac{r_1+r_2-2t}{2t}.$$

Writing $k = 2t$ or $2t+1$ as desired, the probability of obtaining a sequence of r_1 E's and r_2 $\bar{E}$'s arranged in k sets is

$$P\{k \mid r_1, r_2, H_0\} = \frac{p^{r_1} q^{r_2} f_k}{\sum_{\text{All } t} p^{r_1} q^{r_2} f_t} = \frac{f_k}{\sum_{\text{All } t} f_t}.$$

k may take values $2, 3, \ldots, 2r_2$, if $r_1 = r_2$, and values $2, 3, \ldots, 2r_2+1$ if $r_1 > r_2$.

4. Following the orthodox procedure, in order to test the hypothesis, H_0, it is necessary to find two numbers k_1 and k_2 such that

$$P\{k < k_1 \mid H_0\} \leqslant \tfrac{1}{2}\epsilon, \quad P\{k > k_2 \mid H_0\} \leqslant \tfrac{1}{2}\epsilon,$$

and therefore

$$P\{k_1 \leqslant k \leqslant k_2\} \geqslant 1-\epsilon,$$

where ϵ is a number arbitrarily at choice. If an observed number of sets, say k', falls outside the limits k_1 and k_2 then the hypothesis H_0 will be rejected in favour of some alternate hypothesis, H_1. Alternately if H_0 is not true, but H_1 is, then

$$1 - P\{k_1 \leqslant k \leqslant k_2 \mid H_1\}$$

will be the power of the test in the sense of the word as used by Neyman & Pearson. Whether k_1 or k_2 is chosen to judge the significance of an observed k' will depend on which departure from randomness it is most important not to overlook. If the alternate hypothesis is that there is positive dependence in the chain, i.e. that E having occurred in the sth trial it is more likely to occur in the $(s+1)$st trial, then k_1 would be chosen. Such a situation was envisaged in a proposed smooth test to supplement the χ^2 criterion (David, 1947). If, however, the alternate hypothesis is that there is negative dependence, i.e. that E having occurred in the sth trial, it is less likely to occur in the $(s+1)$st trial, then k_2 would be the appropriate criterion. If it is immaterial whether the departure from randomness is positive or negative dependence, then both k_1 and k_2 may be used.

5. We now consider the alternate hypothesis, H_1. Write E_s for the occurrence of the event E in the sth trial and $\bar{E}_s$ for its negation. Let

$$P\{E_1\} = P, \qquad P\{\bar{E}_1\} = Q, \quad P+Q = 1 \text{ and } P \geqslant Q,$$
$$P\{E_s \mid E_{s-1}\} = p_1, \quad P\{\bar{E}_s \mid E_{s-1}\} = q_1,$$
$$P\{E_s \mid \bar{E}_{s-1}\} = p_2, \quad P\{\bar{E}_s \mid \bar{E}_{s-1}\} = q_2.$$

Thus p_1 and q_2 are probabilities of no change and p_2 and q_1 probabilities of a change. If the events are independent then

$$p_1 = p_2 = P \quad \text{and} \quad q_1 = q_2 = Q.$$

6. In calculating the probability of obtaining any given sequence, what will matter will be the number of changes from E to $\bar{E}$ and back again. Let $f_t(r_1)$ be the number of ways in which r_1 E's can be arranged in t groups, i.e. let

$$f_t(r_1) = \frac{(r_1-1)!}{(t-1)!\,(r_1-t)!}.$$

If there are $2t$ groups in a sequence of r_1 E's and r_2 $\bar{E}$'s, the number of ways of obtaining such a sequence will be

$$f_t(r_1)\, f_t(r_2)$$

if the sequence starts with E or with $\bar{E}$. The probability of obtaining any given sequence of r_1 E's and r_2 $\bar{E}$'s of $2t$ groups will be

$$P p_2^{t-1} p_1^{r_1-t} q_1^t q_2^{r_2-t} \quad \text{or} \quad Q q_1^{t-1} q_2^{r_2-t} p_2^t p_1^{r_1-t}.$$

This follows from the fact that a sequence of $2t$ groups beginning with E will imply t changes from E to $\bar{E}$ and $t-1$ changes from $\bar{E}$ to E. The changes are reversed in number if the sequence starts with $\bar{E}$. For $2t+1$ groups the number of ways of obtaining the sequence will be

$$f_{t+1}(r_1)\, f_t(r_2) \quad \text{or} \quad f_t(r_1)\, f_{t+1}(r_2)$$

according as the sequence begins with E or $\bar{E}$. The respective probabilities will be

$$P p_2^t p_1^{r_1-t-1} q_1^t q_2^{r_2-t} \quad \text{and} \quad Q q_1^t q_2^{r_2-t-1} p_2^t p_1^{r_1-t}.$$

The probability therefore of obtaining a sequence of $r_1 E$'s and $r_2 \bar{E}$'s in $2t$ groups will be therefore, under hypothesis H_1,

$$P\{2t \mid r_1 r_2 H_1\} = \frac{\left(\frac{p_2 q_1}{p_1 q_2}\right)^t f_t(r_1) f_t(r_2) \left(\frac{P}{p_2} + \frac{Q}{q_1}\right)}{\sum_{t=1}^{r_2} \left(\frac{p_2 q_1}{p_1 q_2}\right)^t \left[f_t(r_1) f_t(r_2) \left(\frac{P}{p_2} + \frac{Q}{q_1}\right) + \frac{P}{p_1} f_{t+1}(r_1) f_t(r_2) + \frac{Q}{q_2} f_t(r_1) f_{t+1}(r_2) \right]}.$$

The probability of obtaining $r_1 E$'s and $r_2 \bar{E}$'s in $2t+1$ groups will be similarly

$$P\{2t+1 \mid r_1 r_2 H_1\} = \frac{\left(\frac{p_2 q_1}{p_1 q_2}\right)^t \left[\frac{P}{p_1} f_{t+1}(r_1) f_t(r_2) + \frac{Q}{q_2} f_t(r_1) f_{t+1}(r_2) \right]}{\sum_{t=1}^{r_2} \left(\frac{p_2 q_1}{p_1 q_2}\right)^t \left[f_t(r_1) f_t(r_2) \left(\frac{P}{p_2} + \frac{Q}{q_1}\right) + \frac{P}{p_1} f_{t+1}(r_1) f_t(r_2) + \frac{Q}{q_2} f_t(r_1) f_{t+1}(r_2) \right]}.$$

7. So far no mention has been made of any possible connexion between p_1, q_1, p_2 and q_2. It is obvious in all cases we shall have

$$p_1 + q_1 = 1, \quad p_2 + q_2 = 1,$$

but the connexion between p_1 and p_2 is not immediate. We shall make the simplifying assumption which is perhaps most closely related to practical problems, and shall state that where nothing is known about the $s-1$ trials preceding the sth trial, $P\{E_s\} = P$ and $P\{\bar{E}_s\} = Q$. Under this assumption we have

$$p_2 = \frac{P q_1}{Q}, \quad q_2 = 1 - \frac{P q_1}{Q}.$$

This result is reached easily by noticing that

$$P\{E_s\} = P\{E_s E_{s-1}\} + P\{E_s \bar{E}_{s-1}\} = P\{E_{s-1}\} P\{E_s \mid E_{s-1}\} + P\{\bar{E}_{s-1}\} P\{E_s \mid \bar{E}_{s-1}\}$$

whence

$$P = P p_1 + Q p_2.$$

8. The alternative hypothesis chosen to illustrate the power function formulae is that there is positive dependence in the sequence, i.e. k_1 is found so that

$$P\{k \leqslant k_1 \mid H_0\} \leqslant \epsilon \quad \text{and} \quad 1 - P\{k \geqslant k_1 \mid H_1\}$$

is calculated, when $p_1 \geqslant P$. For economy of drawing, several power curves or what are really sections of a kind of power surface, plotted to coordinates P, p_1, have been put together in the diagrams of Fig. 1. For example the bottom left-hand diagram shows for $r_1 = r_2 = 10$ sections of the conditional power surface for $P = 0{\cdot}5$, $0{\cdot}6$ and $0{\cdot}75$. When H_0 is true and $P = p_1$, we have the 5 % risk of rejecting H_0 wrongly. As $p_1 - P$ increases the chance of detecting the f. ct increases, but in a way dependent on P. The other three diagrams show similar sections of the surfaces with $r_1 = r_2 = 5$, with $r_1 = 14, r_2 = 6$ and with $r_1 = 7, r_2 = 3$. In practice it will not be known what the value of P is, but the curves show reasonably well how the power of the test varies as P and p_1 (and therefore p_2) vary. It is clear that the test for randomness under discussion is most powerful when the numbers of alternates are equal, i.e. when $r_1 = r_2$. The power declines sharply when r_1 increases at the expense of r_2. Another point which emerges is that the test is only moderately powerful, against the given alternate hypothesis tested, when $r_1 + r_2 = 20$, and it would appear therefore that if it was desired not to overlook a possible departure from randomness in the form of positive dependence in the chain, then the length of the sequence should consist of at least 20 units. The question of other possible tests we shall not discuss at this stage.

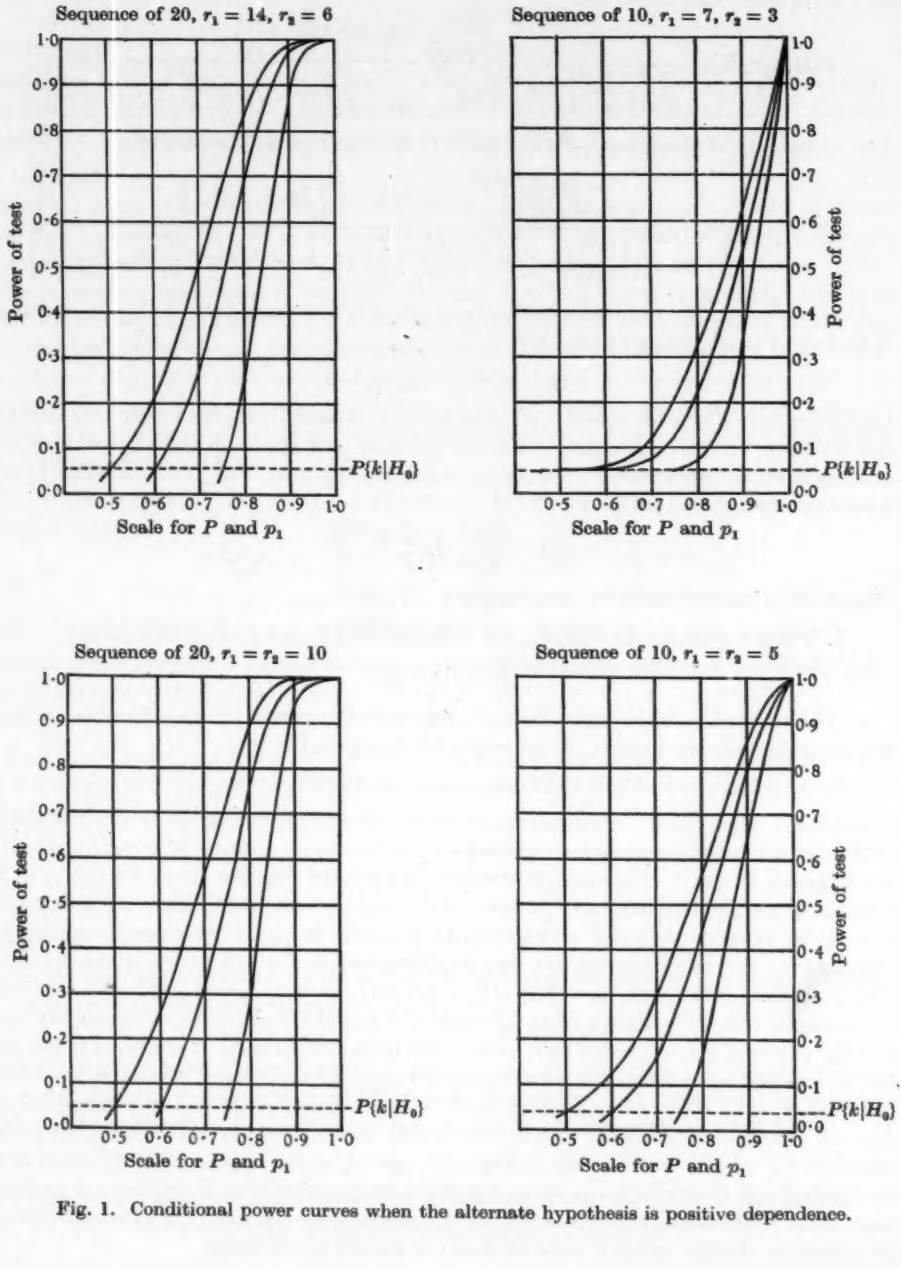


Fig. 1. Conditional power curves when the alternate hypothesis is positive dependence.

9. It will be noticed that $P\{2t \text{ or } 2t+1 \mid r_1 r_2 H_1\}$ which have been loosely termed power function formulae are not power functions in the sense originally defined by Neyman & Pearson, but they appear to involve a justifiable extension of that idea. In order to distinguish them from the usual meaning of the words power function, I shall refer to them as *conditional power functions*. The theory of the conditional power function may be stated briefly in the following way. It is assumed that all possible samples (or sequences) may be classified according to their composition. Suppose that there are k of these mutually exclusive classes, which are also the only possible, say $C_1, C_2, \ldots, C_k$. We have considered only the case where k is finite but it appears likely that the method can be extended to cover the case where k is enumerably infinite. These classes, $C_1, C_2, \ldots, C_k$ will correspond to regions forming a partition of the sample space.

Let H_0 be the hypothesis tested and w_0 be the critical region used for the rejection of this hypothesis. Given that a sample is in C_i (say), and that an alternate hypothesis H_1 is true, then the probability that H_0 will be rejected is

$$P\{E \epsilon w_0 C_i \mid E \epsilon C_i, H_1\} = \frac{P\{E \epsilon w_0 C_i \mid H_1\}}{P\{E \epsilon C_i \mid H_1\}}$$

where $w_0 C_i$ means the region common to w_0 and to C_i and, following the Neyman-Pearson notation, E is the sample point. Regarded as a function of H_1 this is the conditional power function of the test associated with w_0 in the subset C_i of samples.

The Neyman-Pearson power function, which we might call here the *overall* power function, will be

$$P\{E \epsilon w_0 \mid H_1\} = \sum_{i=1}^{k} P\{E \epsilon w_0 C_i \mid H_1\} = \sum_{i=1}^{k} P\{E \epsilon w_0 C_i \mid E \epsilon C_i, H_1\} P\{E \epsilon C_i \mid H_1\},$$

which may be looked on as a weighted average of the conditional power functions.

10. There seems to be no reason why w_0 should not be built up of portions $w_0 C_i$, these portions being chosen to maximize each term of the summation, i.e. $w_0 C_i$ chosen to maximize the conditional power function. For example, to revert to the specific case of randomness within a sequence with which we have been dealing, the different partitions of $r\ (= r_1 + r_2)$ are the mutually exclusive and only possible classes C_i. It is conceivable, although practically not very likely, that for each of these classes there will exist a different test which is more powerful to detect specifically defined departures from the basic hypothesis tested than any other test. The decision as to which is the most powerful test, against the same specifically defined alternatives, to use for any given class will be decided by the conditional power function. Once this has been decided the procedure for the complete test of significance may be laid down. This will be: (i) count the number of alternatives in the sequence, i.e. find r_1 and r_2, (ii) from (i) decide the appropriate test of significance to use, (iii) apply the test. The power of the test as laid down by (i), (ii) and (iii), in the usual meaning of the word, will be given by the overall power function.

It is proposed to discuss these, and other applications of the conditional power function technique, in a further publication. I have been concerned here with trying to explain what I believe to be the basic ideas, and to forestall possible criticism that I am falling into error (of the third kind) and am choosing the test falsely to suit the significance of the sample.

REFERENCES

STEVENS, W. L. (1939). *Ann. Eugen., Lond.*, **9**, 11.
WALD, A. & WOLFOWITZ, J. (1940). *Ann. Math. Statist.* **11**, 147.
DAVID, F. N. (1947). *Biometrika*, **34**, 299.

A NUMERICAL SOLUTION OF THE PROBLEM OF MOMENTS

By H. O. HARTLEY and S. H. KHAMIS

1. Introduction

Given a statistical variable x and its frequency distribution $f(x)$, then, under certain continuity conditions for $f(x)$, the moments

$$\mu_r = \int x^r f(x)\,dx \quad (r = 0, 1, 2, \ldots) \tag{1}$$

can be evaluated for any integer r. For certain distributions $f(x)$ the integrations in (1) can be carried out analytically resulting in simple formulae for the moments. In general there is no inherent difficulty in obtaining numerical values for the moments by numerical quadrature.

The inverse problem is to find the distribution $f(x)$ given the moments μ_r. This problem, commonly known as 'The Problem of Moments', has received considerable attention by mathematicians and is of interest in statistical distribution theory. There are numerous statistics for which it is difficult to obtain a formula of the random sampling distribution $f(x)$ amenable to numerical evaluation. On the other hand, in such cases it is often possible to find simple formulae for the random sampling moments (Bartlett, 1937). Sometimes such formulae are available for *all* integer r; more often than not, however, μ_r is only known for a limited number of small r (e.g. $r = 0, 1, \ldots, 6$). A simple method of 'determining' $f(x)$ from the given moments would therefore be helpful in such cases.

Examples of variables of this kind are the numerous moment statistics or k-statistics for which random sampling moments can be evaluated, notably by R. A. Fisher's (1929, 1930) combinatorial methods, whilst their exact sampling distributions are usually unknown. As related statistics we should mention here the moment ratios $\sqrt{b_1}$ and b_2 used in tests for deviation from normality (Geary, 1947, Geary & Worledge, 1947). For these, the low-order moments are known exactly. A similar situation arises with statistics defined as likelihood ratios, as, for instance, with the criterion L_1 required for testing heterogeneity in a set of variances. Moments for this statistic were obtained by Neyman & Pearson as early as 1931, yet, although approximations to $f(L_1)$ have been obtained (Bartlett, 1937; Hartley, 1940; Nayer, 1936; Neyman & Pearson, 1931; Sukhatme, 1936; Welch, 1935, 1936), there is still considerable doubt about their accuracy in certain cases, and the exact formula obtained by Nair (1936) in the case of equal sample sizes is very complex.

These and numerous other problems of distribution point to the necessity of developing a numerical technique to deal with the following situation:

(i) A random variable x ranging between a and b (where a may be $-\infty$ and b may be $+\infty$) has a distribution function $f(x)$ known to have a continuous derivative of order n.

(ii) The moments

$$\mu_r = \int_a^b x^r f(x)\,dx \quad (r = 0, 1, \ldots, R), \tag{2}$$

are known numerically to any decimal accuracy desired but for a limited number of positive integers r, viz. $r = 0, 1, \ldots, R$. With the knowledge about $f(x)$ limited to the above conditions, is it possible to obtain numerical values for the probability integral $P(x) = \int_a^x f(x)\,dx$

depending on the moments only, and is it possible to make a statement on the accuracy of these values in terms of the derivatives of the function $f(x)$?

Problems of this kind have hitherto been treated principally in two ways:

(*a*) When $R = 2, 3$ or 4 nothing better can be expected than a 'good fit', which is often achieved by fitting the appropriate Pearson-type curve.

(*b*) With R in the neighbourhood of 5–8, expansions of the Gram Charlier, Laguerre or Jacobi type have been used, either as cumulant or as moment expansions. Such theorems as are available for statements on the convergence and asymptotic behaviour of these expansions usually require too many moments to be known. Often the expansions are only asymptotic, and unless the distribution is close to the generating curve (Normal for Gram Charlier, Γ for Laguerre), the results are often disappointing (see, for example, Kendall, 1945, Chapter 6).

2. OUTLINE OF PRESENT METHOD

The method to be developed here is a direct application of finite-difference calculus and therefore provides both numerical answers to the problem, as well as gauges of their accuracy in form of remainder terms. The method is, in fact, closely linked with interpolation technique. When using any of the well-known interpolation formulae no mathematically rigorous statement on the accuracy of the interpolates can be made unless the magnitude of the remainder term can be estimated, and for this some knowledge about (say) the nth derivative of the function is required. Yet, in using such formulae the convergence of the difference table inspires confidence that 'the results of the interpolation can be accepted as a working hypothesis' (Milne Thomson, 1933, p. 62). Similarly, with the present method we shall give a numerical procedure of obtaining values of the probability integral. Certain checks of internal consistency will be described which inspire confidence that the answers are correct, but no rigorous statement on the accuracy can, of course, be made if this is to be based on a finite number of moments alone. The *exact* remainder terms which we derive will entail the high-order derivatives of $f(x)$, and it is hoped, in a second communication, to derive some *general* statements concerning their order of magnitude.

In order to simplify the argument we assume in this section that the range of x is finite (a and b finite).

The aim is to determine the probability integral of x, $P(x) = \int_a^x f(x)\,dx$ in tabular form, i.e. we wish to determine numerical values of

$$P_i = P(x_i) = \int_a^{x_i} f(x)\,dx \tag{3}$$

for discrete values of x_i. For convenience the group intervals $x_{i+1} - x_i$ will generally be chosen equidistant (group interval $= h$), and the number of intervals will be $R+1$, i.e. equal to the number of given moments (including $\mu_0 = 1$). Hence

$$x_i = a + ih, \quad h = (b-a)/(R+1). \tag{4}$$

The first differences in the table derived from equation (3) are the quantities

$$f_i = P_i - P_{i-1} = \int_{x_{i-1}}^{x_i} f(x)\,dx, \tag{5}$$

and are the familiar 'frequencies' f_i in a grouped frequency distribution with equidistant intervals (see Fig. 1). The link between these frequencies and the exact moments μ_r is then

established by the well-known formulae for Sheppard's correction. Using Kendall's (1938, 1945) derivation and remainder term, but extending his notation, we have

$$\sum_{i=1}^{R+1} f_i \xi_i^r = \mu_r + C(r,h) + S(r,h), \tag{6}$$

where the centre points ξ_i are given by

$$\xi_i = a + (i - \tfrac{1}{2})h \quad (i = 1, \ldots, R+1), \tag{7}$$

$C(r,h)$ denotes Sheppard's corrective term, viz.

$$C(r,h) = \sum_{j=1}^{[\frac{1}{2}r]} \left(\frac{h}{2}\right)^{2j} \binom{r}{2j} \frac{1}{2j+1} \mu_{r-2j}, \tag{8}$$

and $S(r,h)$ the remainder term.

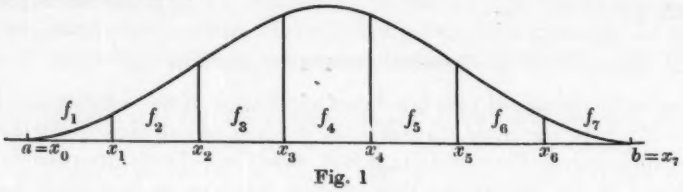


Fig. 1

The aim, now, is to use equations (6) to determine the unknown f_i from the given μ_r. To this end the remainder $S(r,h)$ must be examined: Most distribution functions have what is commonly known as high contact at the terminals of the variate range. This means that $f(x)$, as well as all its derivatives up to order, say, m, vanish at both ends of the range, i.e.

$$f^{(i)}(a) = f^{(i)}(b) = 0 \quad (i = 0, \ldots, m). \tag{9}$$

If for such functions we define $f(x) = 0$ outside the range $a \leqslant x \leqslant b$, it will have continuous derivatives of up to order m for $-\infty < x < +\infty$. It can then be shown that the remainder term is of the form (see, for example, Kendall, 1945, p. 69)

$$S_m(r,h) = -\frac{(R+1)h^m}{m!} B_m k^{(m)}(r,h,\theta_r) \quad (m \text{ even}), \tag{10}$$

$$S_m(r,h) = \frac{2(R+1)h^m}{m!} B_{m+1}^{(1)}(\tfrac{1}{2}) k^{(m)}(r,h,\theta_r) \quad (m \text{ odd}), \tag{11}$$

$$a \leqslant \theta_r \leqslant b,$$

where the B_j are the Bernoulli numbers, the $B_j^{(1)}$ are the Bernoulli polynomials of first order, the integrand function $k(r,h,x)$ is defined by

$$k(r,h,x) = x^r \int_{-\frac{1}{2}h}^{+\frac{1}{2}h} f(x+\xi)\, d\xi, \tag{12}$$

and its derivatives with regard to x are denoted by $k^{(i)}$. In the subsequent sections we shall assume (9) to hold (contact of order m), but will discuss the case when (9) is not satisfied in § 10.

The remainder term $S_m(r,h)$ will usually be small (see, for example, Kendall, 1945, p. 72). We shall therefore, in what follows, ignore $S_m(r,h)$ but will discuss the error thereby committed in § 5.

If, then, in (6) we omit $S(r,h)$ we obtain a system of $R+1$ linear equations for the $R+1$ unknowns f_i

$$\sum_{i=1}^{R+1} f_i \xi_i^r = \mu_r + C(r,h) = \overline{\mu}_r. \tag{13}$$

The matrix of this system of equations (v_R say) is of the form $|\xi_i^r|$ and has a classical determinant $\|\xi_i^r\|$, sometimes referred to as Vandermonde's determinant and well known to be $\neq 0$. The system can therefore be inverted once and for all and, for any *particular* case, the unknown f_i can then be determined by substituting the right-hand sides of (13), i.e. $\overline{\mu}_r$ in the inverse matrix v_R^{-1}. Denoting the elements of this inverse matrix by u_{ir} we have the system of equations

$$f_i = \sum_{r=0}^{R} u_{ir}\overline{\mu}_r. \tag{14}$$

Progressive addition of the f_i yields the P_j from $P_j = \sum_{i=1}^{j} f_i$* and therefore a table of $P(x)$ at interval h: Finally, intermediate values of $P(x)$ can be obtained by standard interpolation. Alternatively, as described in § 7, we may obtain directly a table of $P(x)$ at interval $\frac{1}{2}h$.

3. THE STANDARD FORM OF THE NUMERICAL INVERSION

The rank of the original matrix v_R is obviously equal to $R+1$, i.e. the number of moments given, whilst its elements are the powers of the centre points ξ_i^r. It is desirable therefore that, for any given R, scale and location of the variable x be transformed into a standard form X, so that only *one* matrix V_R and therefore only *one* matrix V_R^{-1} need be calculated for each R. It is most convenient to standardize as follows:

$$X = (x - \tfrac{1}{2}(a+b))\frac{R+1}{b-a} \quad (R \text{ even}), \tag{15}$$

$$X = (x - \tfrac{1}{2}(a+b))\frac{R+1}{b-a} + \frac{1}{2} \quad (R \text{ odd}). \tag{16}$$

It will be seen, therefore, that the range of X is $R+1$ and the group interval

$$H = X_{i+1} - X_i = 1.$$

From the given moments of x those of X (M_r say) about $X = 0$ can, of course, be calculated by the usual binomial formulae, and in what follows we assume that values of M_r are given numerically. Further, in analogy to (13), we have

$$\overline{M}_r = M_r + C(r, 1). \tag{17}$$

From (15) and (16) we obtain for the new centre points

$$\left.\begin{aligned} \Xi_i &= -\tfrac{1}{2}R, \ldots, 0, \ldots, +\tfrac{1}{2}R \quad \text{for even } R, \\ \Xi_i &= -\frac{R-1}{2}, \ldots, 0, \ldots, \frac{R+1}{2} \quad \text{for odd } R, \end{aligned}\right\} \tag{18}$$

and the matrix V_R becomes $|(i-1-\frac{1}{2}R)^r|$ or $|(i-\frac{1}{2}R-\frac{1}{2})^r|$. Thus, if the first six moments are given, we obtain for V_6:

$$V_6 \equiv \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \\ -27 & -8 & -1 & 0 & 1 & 8 & 27 \\ 81 & 16 & 1 & 0 & 1 & 16 & 81 \\ -243 & -32 & -1 & 0 & 1 & 32 & 243 \\ 729 & 64 & 1 & 0 & 1 & 64 & 729 \end{vmatrix} \tag{19}$$

* It is, of course, possible to construct a matrix yielding the P_j directly from the $\overline{\mu}_r$, but we are here satisfied with determining the f_i first, as they are of independent interest.

In practice the important range of R will be from 5 to about 8. The inverse matrix V_6^{-1} is given below, and it is hoped to give V_7^{-1}, V_8^{-1} and V_5^{-1} in a subsequent paper. The inverse matrix V_6^{-1}, the elements of which are denoted by U_{ir}, can be written in the form

$$c_i f_i = \sum_{r=0}^{R} U'_{ir} \bar{M}_r, \tag{20}$$

where $U'_{ir} = c_i U_{ir}$, i.e. the c_i are suitable common denominators of the $U_{i0}, \ldots, U_{iR}$, and the U'_{ir} are given in the body of the schedule below:

i	$c_i f_i$	$\bar{M}_0 = 1$ $r=0$	$\bar{M}_1$ 1	$\bar{M}_2$ 2	$\bar{M}_3$ 3	$\bar{M}_4$ 4	$\bar{M}_5$ 5	$\bar{M}_6$ 6	= multiplier of column
1	$720f_1$	0	−12	4	15	−5	−3	1	
2	$120f_2$	0	18	−9	−20	10	2	−1	
3	$48f_3$	0	−36	36	13	−13	−1	1	
4	$36f_4$	36	0	−49	0	14	0	−1	
5	$48f_5$	0	36	−36	−13	−13	1	1	
6	$120f_6$	0	−18	−9	20	10	−2	−1	
7	$720f_7$	0	12	4	−15	−5	3	1	

(21)

In order to use the above system of equations it would be necessary to compute the $\bar{M}_r$ from the given M_r, using formula (17). It is obviously more convenient to evaluate, once and for all, a matrix U''_{ir} giving the f_i directly in terms of the given M_r. This matrix is given below for $R = 6$:

i	f_i	$M_0 = 1$	M_1	M_2	M_3	M_4	M_5	M_6
1	f_1	0·000 379	−0·011 719	0·002 344	0·017 361	−0·005 208	−0·004 167	0·001 389
2	f_2	−0·005 227	0·109 375	−0·034 896	−0·152 778	0·072 917	0·016 667	−0·008 333
3	f_3	0·059 161	−0·683 594	0·618 490	0·253 472	−0·244 792	−0·020 833	0·020 833
4	f_4	0·891 373	0	−1·171 875	0	0·354 167	0	−0·027 778
5	f_5	0·059 161	0·683 594	0·618 490	−0·253 472	−0·244 792	0·020 833	0·020 833
6	f_6	−0·005 227	−0·109 375	−0·034 896	0·152 778	0·072 917	−0·016 667	−0·008 333
7	f_7	0·000 379	0·011 719	0·002 344	−0·017 361	−0·005 208	0·004 167	0·001 389

(22)

Working rule: Each f_i is obtained by forming the sum of seven products using the seven coefficients in the ith line and applying them to $M_0, \ldots, M_6$, e.g. $f_1 = 0{\cdot}000\,379M_0 - 0{\cdot}011\,719M_1 + \ldots + 0{\cdot}001\,389M_6$.

4. Calculation of the incomplete B-function $I_x(8, 6)$ from its first six moments

As an example for the above method we consider the Beta Distribution for $p = 8$ and $q = 6$, viz.

$$f(x) = [B(8,6)]^{-1} x^7 (1-x)^5.$$

Using the moments for this distribution about $x = 0$, $\mu_r = B(x+r, 6)/B(8, 6)$ $(r = 0, \ldots, 6)$ and transforming to the standard scale $X = 7x - 3{\cdot}5$, we obtain for the moments of X about $X = 0$: $M_1 = 0{\cdot}5$, $M_2 = 1{\cdot}05$, $M_3 = 1{\cdot}225$, $M_4 = 2{\cdot}77426$, $M_5 = 4{\cdot}41360$ and $M_6 = 10{\cdot}56942$. Substituting these in the matrix (22) we obtain values of f_i whose progressive sums are shown in Table 1 (calculated $I_x(8, 6)$). These may be compared with the 'exact' values obtained (by interpolation) from the *Tables of the Incomplete B-function* (1934). The worst discrepancy is about 2 in the fourth decimal. Higher accuracy can, of course, be obtained if the number of moments $(R+1)$ and therefore the number of f_i increases (see, for example, § 8, where the normal curve is obtained to 5-decimal accuracy).

A rather gratifying feature of the comparison is the higher *decimal* accuracy in the tails of the distribution. This is a consequence of the sensitivity of the higher moments to changes in the tail frequencies. Note also that the elements in the top and bottom lines of the inverse matrix (22) are much smaller than those in the other lines, so that any error in the right-hand sides of (13) has a smaller effect on the terminal f_i.

Table 1. *Comparison of 'calculated' and 'exact' values of* $I_x(8, 6)$

X	x	Exact I_x	Calculated I_x	Difference 10^{-5}
−2·5	1/7	0·000 11	0·000 09	2
−1·5	2/7	0·013 41	0·013 54	−13
−0·5	3/7	0·140 17	0·139 95	22
0·5	4/7	0·489 63	0·489 81	−18
1·5	5/7	0·862 61	0·862 70	− 9
2·5	6/7	0·994 11	0·993 95	16
3·5	7/7	1·000 00	1·000 00	0

It might be argued that a further error will arise when determining intermediate values of I_x by interpolation in the 'calculated' table. This difficulty could, however, be overcome by shifting the grid of group intervals and using a standard X-scale with group end-points corresponding to the odd multiples of 1/14 in x, thereby obtaining I_x at points half-way between the arguments of Table 1. Such a method has actually been used in § 7.

5. The remainder term

A formal representation of the remainder term is immediately obtained by reverting to the exact equations (6). If we are concerned with distribution functions having contact of order m at the terminals, the error contributions to the f_i are obtained by substituting the $R+1$ remainder terms $S_m(r, h)$ ((10), (11)) in the inverse matrix v^{-1}. It is convenient to use the standard variate X-scale, $H = 1$ and the V^{-1} matrix when it will be found that

$$\text{error } f_i = \sum_{r=0}^{R} U_{ir} S_m(r, 1), \tag{23}$$

where $S_m(r, 1)$ is given by (10) or (11) putting $h = 1$ and remembering that the integrand function k must be taken in terms of the standard variate X, viz.

$$k(r, 1, X) = X^r \int_{-\frac{1}{2}}^{+\frac{1}{2}} f\left(\frac{b-a}{R+1}(X+\xi)\right)\frac{b-a}{R+1}\,d\xi. \tag{24}$$

Since the arguments θ_r of $k^{(m)}(r, 1, X)$ are unknown it will as a rule be necessary to substitute their respective maxima in (23), at the same time taking $|U_{ir}|$ in place of U_{ir}.

Although with (23) we have given a formal solution of the error term involved, in a manner similar to the remainder terms of interpolation formulae, it will in practice be difficult to estimate the magnitude of the error from this formula. It is hoped, therefore, to go into this aspect more fully in a second paper.

6. Infinite variate range and artificial truncation

When the range of the variate is infinite, i.e. when $a = -\infty$ and/or $b = +\infty$, it is, of course, possible to transform the variate x by, say, $y = y(x)$ such that the range of y is finite. However, in general, we shall not be able to assume that the moments of y are known or that they can

be derived from those of x. It is therefore necessary to adapt our method to deal with an infinite variate range. We shall treat here the case $b = +\infty$, the case $a = -\infty$ being identical and the case $a = -\infty$ *and* $b = +\infty$ being analogous.

For an infinite variate range, the condition of high contact is now replaced by

$$\lim_{x \to \infty} f^{(i)}(x) = 0 \quad (i = 0, 1, \ldots, m), \tag{25}$$

which results in remainder terms analogous to (10) and (11)*. Similarly, in equations (6) which correspond to Kendall's (1945) equations (3·40), the summation now extends from $i = 1$ to $i = \infty$, there being an infinity of frequencies $f_i = \int_{x_{i-1}}^{x_i} f(x)\,dx$. Now since the μ_r exist we know that

$$\int_a^\infty x^r f(x)\,dx \tag{26}$$

is convergent. Accordingly

$$\lim_{b \to \infty} \sum_{i=R+2}^{\infty} (i - \tfrac{1}{2})^r h^r \int_{(i-1)h}^{ih} f(x)\,dx = 0, \tag{27}$$

if $h = (b-a)/(R+1)$. If, therefore, we denote the above sums by $\epsilon(r, b)$ respectively we have, from (6),

$$\sum_{i=1}^{R+1} f_i \xi_i^r + \epsilon(r, b) = \mu_r + C(r, h) + S(r, h). \tag{28}$$

Applying now the previous method we introduce an additional error in the calculation of f_i, but this error is smaller than $\quad + \max |\epsilon(r, b)| \Sigma |u_{ir}|$.

The precise determination of the $\epsilon(r, b)$ for any given b would, of course, require a knowledge of the nature of the convergence in (26), i.e. some external knowledge about the distribution $f(x)$ which we are seeking to determine numerically. Unfortunately, such knowledge will in general not be available.

However, if b is chosen sufficiently large, the f_i determined for *different* values of b should all yield, by the method of §§ 2 and 3, approximations to the *same* probability integral $P(x)$ to within the errors of the respective remainder terms $S(r, h)$ and to within the errors introduced by (27). In practice, therefore, one would make an intelligent guess at the likely range of b and then test for internal consistency by comparing the probability integral tables obtained by varying b over this range. This method, which is illustrated in § 7 gives an idea of the accuracy to which the integral has been determined, but no rigorous statement on accuracy can be made without appealing to some *a priori* knowledge about $f(x)$. It is hoped to deal with this aspect more fully in the next paper.

7. The calculation of the χ-distribution for 10 degrees of freedom

As an illustration of the preceding section, we will now calculate the χ-distribution for 10 degrees of freedom. This distribution has high contact at either terminal and, although it is known to start at $x = \chi = 0$, we shall treat it as a distribution of double infinite range, i.e. we shall not make direct use of the information that $f(x) = 0$ for $x \leqslant 0$, and choose a *truncated* range $a \leqslant x \leqslant b$.

We have a mean of $\mu_1 = 3{\cdot}0843\,2776$, and the moments about the mean are given by†

$$\mu_2' = 0{\cdot}486\,9223, \quad \mu_3' = 0{\cdot}080\,6720, \quad \mu_4' = 0{\cdot}713\,2999, \quad \mu_5' = 0{\cdot}386\,6784, \quad \mu_6' = 1{\cdot}810\,4865.$$

* A formula for $S(r, h)$ when the range is infinite will be given in the second paper.

† These follow from the formulae for the moments about the origin which are ratios of Γ-functions (see, for example, Kendall, 1945, p. 55). Note that we have used μ and μ' for moments about the origin and the mean, respectively.

The standard deviation is $\sqrt{\mu_2'} = 0{\cdot}7$, and with seven group intervals available to cover the essential range we should choose h of the order of the standard deviation.* Our first attempt is, therefore, (*a*) $h = 0{\cdot}8$.

(*a*) If we make the mean of x the centre point of the innermost interval we have for the truncated range $a = \mu_1 - 3{\cdot}5 \times 0{\cdot}8 = \mu_1 - 2{\cdot}8$ and $b = \mu_1 + 2{\cdot}8$. For the standard variate X, the origin $X = 0$ will coincide with the mean of x and its range will be $-3{\cdot}5 \leqslant X \leqslant +3{\cdot}5$. Calculation of the moments (M_r) of X and substitution in the matrix (22) yields the following answers for the frequencies f_i:

$$f_1 = 0{\cdot}000\,5, \quad f_2 = 0{\cdot}033\,25, \quad f_3 = 0{\cdot}262\,66, \quad f_4 = 0{\cdot}424\,71,$$

$$f_5 = 0{\cdot}231\,96, \quad f_6 = 0{\cdot}042\,06, \quad f_7 = 0{\cdot}004\,87.$$

The calculated frequency (f_7) for the interval $\mu_1 + 2{\cdot}0 \leqslant x \leqslant \mu_1 + 2{\cdot}8$ is about 0·005, and its contribution to μ_6' about $0{\cdot}005\,(2{\cdot}4)^6 \sim 1$. Since this is an appreciable proportion of μ_6' it is unlikely that the frequencies beyond $b = \mu_1 + 2{\cdot}8$ when substituted in (27) can be neglected, i.e. b and h are too small.†

(*b*) Choosing therefore a larger h, we try $h = 1$. If we still keep the mean in the centre of the truncated range we have $a = \mu_1 - 3{\cdot}5 = -0{\cdot}42$ and $b = \mu_1 + 3{\cdot}5 = 6{\cdot}58$ (we know, of course, that $f(x) = 0$ for $x = 0$ so that our f_1 will really be the frequency for the interval $0 \leqslant x \leqslant 0{\cdot}085$). This time the standard variate is $X = x - \mu_1$ so that $M_r = \mu_r'$, and the above values can be substituted directly in the matrix (22) yielding the comparison of calculated χ-integral and 'exact' χ-integral as shown in Table 2.

Table 2. *Comparison of calculated and exact values of the χ-integral*

$X = \chi - \mu_1$	$P(x)$ exact	$P(x)$ calculated	Difference 10^{-5}
−2·5	0·000 06	0·000 11	− 5
−1·5	0·009 29	0·008 93	36
−0·5	0·244 66	0·244 75	− 9
+0·5	0·767 67	0·767 85	−18
+1·5	0·979 02	0·978 88	14
+2·5	0·999 45	0·999 47	− 2
+3·5	1·000 00	1·000 00	0

The maximum error is about 0·0004 and, again, the terminal f_i have a higher decimal accuracy. In practice, of course, the exact distribution would not be available for comparison. This time the terminal value f_7 is about 0·0005 and represents the frequency for the interval $\mu_1 + 2{\cdot}5 \leqslant x \leqslant \mu_1 + 3{\cdot}5$. Its contribution to μ_6' is about 0·4, thereby confirming that the previous grid of group intervals was too fine. To obtain further confirmation on the tail of the distribution, we determine a third set of f_i by shifting the grid of group intervals by 0·5 to the right, retaining the interval $h = 1$. This will make $a = \mu_1 - 3$ and $b = \mu_1 + 4$, i.e. $0{\cdot}08 \leqslant x \leqslant 7{\cdot}08$. For our standard variate X the origin will now coincide with $\mu_1 + 0{\cdot}5$.

* An unsuitable choice of h would, later, fail to satisfy the checks of internal consistency.

† Comparison with the exact χ-distribution shows that the maximum error in the above f_i is nevertheless not more than 0·005.

The values of the M_r are as follows:

$$M_1 = -0{\cdot}5, \quad M_2 = 0{\cdot}736\,9223, \quad M_3 = -0{\cdot}774\,7114,$$
$$M_4 = +1{\cdot}344\,8394, \quad M_5 = -1{\cdot}834\,794, \quad M_6 = 3{\cdot}595\,7606.$$

Substituting these in the matrix (22) we obtain the following values of f_i:

$$f_1 = 0{\cdot}000\,15, \quad f_2 = 0{\cdot}070\,12, \quad f_3 = 0{\cdot}444\,30, \quad f_4 = 0{\cdot}404\,21,$$
$$f_5 = 0{\cdot}076\,99, \quad f_6 = 0{\cdot}004\,19, \quad f_7 = 0{\cdot}000\,04.$$

The comparison of the progressive sums of the above f_i with the exact χ-integral is of similar accuracy to that in Table 2. The terminal frequency for $\mu_1+3 \leqslant x \leqslant \mu_1+4$ is 0·0005 with a contribution of about 0·03 to μ_6', indicating that we have now reached a satisfactory choice of b.

As a final check on the internal consistency we compare the answers obtained with the two last choices of group intervals by merging the tables of $P(x)$ to obtain one table at interval 0·5. This is set out in Table 3. The differences provide a fair check on the internal consistency to about 3-decimal accuracy of the two separate tables. If a more reliable check is desired, three or even four separate tables may be computed, all at the same group interval h and merged in the above manner to form a single table at interval $\frac{1}{3}h$ or $\frac{1}{4}h$. This procedure has the added advantage that interpolation difficulties at the wide interval of h are being avoided.

Table 3. *$x = \chi$ for 10 degrees of freedom. Calculated table of $P(x)$ obtained from two separate grids of group intervals ($h = 1$)*

$x-\mu_1$	$P(x)$			
−2·5	0·0001			
		1		
−2·0	0·0002		86	
		87		442
−1·5	0·0089		528	
		615		601
−1·0	0·0704		1129	
		1744		− 174
−0·5	0·2448		955	
		2699		−1122
0	0·5147		− 167	
		2532		− 855
0·5	0·7679		−1022	
		1510		112
1·0	0·9189		− 910	
		600		480
1·5	0·9789		− 430	
		170		296
2·0	0·9959		− 134	
		36		103
2·5	0·9995		− 31	
		5		
3·0	1·0000			

8. The special case of symmetrical distributions; the normal integral

By placing the origin of the standard variate X at the mean of a symmetrical distribution we obviously have $f_1 = f_{R+1}$, $f_2 = f_R$, etc., i.e. the number of unknowns is halved. On the other hand, the odd moments contribute the meaningless equations

$$\Sigma f_i(\Xi_i^r + \Xi_{R+2-i}^r) = \Sigma f_i \times 0 = 0.$$

With the number of unknowns and equations halved and with even moments only retained, it is necessary to work out a new matrix ($\hat{V}_R$ say) based on even-order moments only. In practice the important values of R are $R = 4, 6, 8$ and 10, and we are giving below the inverse matrix $\hat{V}_8^{-1}$ (for $R = 8$) having rank 5 (as there are five equations corresponding to μ_0, μ_2, μ_4, μ_6 and μ_8):

i	f_i	$M_0 = 1$	M_2	M_4	M_6	M_8
1	f_1	0·000 3441	−0·001 7857	0·001 2153	−0·000 2315	0·000 0124
2	f_2	−0·003 9874	0·020 8333	−0·013 7153	0·002 3148	−0·000 0868
3	f_3	0·022 4151	−0·119 0476	0·071 1806	−0·008 1019	0·000 2480
4	f_4	−0·088 4281	0·500 0000	−0·143 4028	0·012 7315	−0·000 3472
5	f_5	0·569 6563	−0·400 0000	0·084 7222	−0·006 7130	0·000 1736

(29)

Working rule: Each f_i is obtained by forming the sums of five products using the five coefficients in the ith line and applying them to $M_0, \ldots, M_8$; e.g. $f_1 = 0{\cdot}000\,3441 M_0 - \ldots + 0{\cdot}000\,0124 M_8$.

As an example we compute the normal integral from its first five even moments, μ_0 to μ_8, choosing $h = 1$ and the standard variate X as normal deviate. Substituting, therefore, in the matrix (29) $M_0 = 1$, $M_2 = 1$, $M_4 = 3$, $M_6 = 15$ and $M_8 = 105$, we obtain the five f_i which in Table 4 have been progressively added to form the 'calculated normal integral' to be compared with the 'exact' one. The accuracy is remarkable, the maximum error being 15 in the 6th decimal.

Table 4. *Comparison of calculated normal integral with exact normal integral*

$x = X$	Exact $P(x)$	Calculated $P(x)$	Difference $\times 10^{-6}$
−4	0·000 032	0·000 034	− 2
−3	0·001 350	0·001 342	8
−2	0·022 750	0·022 765	−15
−1	0·158 655	0·158 643	12
0	0·500 000	0·500 000	0

With symmetrical distributions we cannot, of course, shift the grid of group intervals, as otherwise we would lose the symmetry relation between the f_i. If, therefore, intermediate values of $P(x)$ are required in order to ease subsequent interpolation, we can achieve this only by altering h. Merging the answers obtained from (say) three different h grids all centred at $x = 0$ (e.g. $h = 0{\cdot}9$, $1{\cdot}0$ and $1{\cdot}1$), we would *not* obtain a table of $P(x)$ at an *equidistant* interval. In the internal check we would, therefore, use divided differences.

9. DIVERGENT OR POORLY CONVERGENT MOMENTS; THE t-DISTRIBUTION FOR 10 DEGREES OF FREEDOM

Some variates with infinite range have distribution functions with low contact at $x = \infty$, i.e. the convergence in

$$\lim_{x \to \infty} f(x) = 0 \,. \tag{30}$$

is slow, indeed, in some cases the moment μ_r is divergent for, say, $r \geqslant R'$.

As an example we have investigated the t-distribution for 10 degrees of freedom. Here we have $f(x) = c(1 + t^2/10)^{-5{\cdot}5}$ and hence $R' = 10$. In this case, therefore, R' is known *a priori*. If no such mathematical information is available, warning of low contact is given by the rapid

growth of the moments as $r \to R$, provided R is near to R'.* For our example for the t-distribution we find

$$\mu_2 = 1{\cdot}25, \quad \mu_4 = 6{\cdot}25, \quad \mu_6 = 78{\cdot}125, \quad \mu_8 = 2734{\cdot}375.$$

The difficulty with such distributions is that artificial truncation is not justified if the high-order, poorly convergent moments are to be used in equations (6). The remedy in such cases is the square variate transformation $y^2 = x$. Sometimes it may be necessary to use a higher power $y^k = x$. Obviously, if we were to take an equidistant interval for y, the group integral for x will *grow* with the square law, thereby absorbing the slowly convergent tail end of $f(x)$.

Now, obviously, the moments of y are simply related to those of x; we have

$$\int_0^\infty x^r f(x)\,dx = 2\int_0^\infty y^{2r} y f(y^2)\,dy, \tag{31}$$

or introducing the new distribution function $g(y) = 2yf(y^2)$, we have

$$\int_0^\infty x^r f(x)\,dx = \int_0^\infty y^{2r} g(y)\,dy. \tag{32}$$

Applying now the previous method to $g(y)$ it is further necessary to avoid using the poorly convergent high-order moments. In the case of the t-distribution, instead of taking $r = 0, 2, 4, 6$ and 8, we take the absolute moments† for $r = 0, 1, 2, 3$ and 4, which, according to (32), correspond to the even moment of $g(y)$. If only even moments about the origin are used in the determination of the f_i, the matrix (29) gives the appropriate inversion. Using $h = 0{\cdot}6$ for the y-group interval we substitute in (29):

$M_0 = 1$, $M_2 = 2{\cdot}401\,906$, $M_4 = 9{\cdot}645\,062$, $M_6 = 52{\cdot}952\,032$ and $M_8 = 372{\cdot}108\,863$.

We thereby obtain five values of f_i ($i = 1, \ldots, 5$) of the form

$$f_i = \int_{(5-i)h}^{(6-i)h} g(y)\,dy = \int_{(5-i)^2h^2}^{(6-i)^2h^2} f(x)\,dx. \tag{33}$$

The progressive sums of these are compared with the corresponding values of the exact t-integral in Table 5. Although the accuracy is lower than in the previous example it is satisfactory and very much better than we could have obtained without applying the transformation $y^2 = x$.

Table 5. *Comparison of calculated and exact values of the t-integral*

t	$P(t)$ exact	$P(t)$ calculated	Difference $\times 10^{-4}$
5·76	0·0001	0·0001	0
3·24	0·0044	0·0042	2
1·44	0·0902	0·0905	− 3
0·36	0·3613	0·3648	− 35
0·00	0·5000	0·5000	0

* If R is much smaller than R', the present difficulty will not arise at all.

† We shall show in a second paper that, if the absolute moments of a distribution are not known, they can be obtained by interpolation between the values of $\log \mu_r$ for $r = 2, 4, 6, 8$, etc.; in fact, we shall give a general discussion of the interpolability of the logarithmic moment function for positive x.

10. LACK OF HIGH CONTACT AT THE START OF THE VARIATE $x=a$

We confine ourselves here to the most important case of lack of high contact at one terminal, say the start of the distribution, and assume, therefore, that there *is* high contact at *one end of the range*.

Without loss of generality we assume that $a = 0$, i.e. $x \geqslant 0$, and introduce the new variate $y^k = x$, $k \geqslant 2$. Whence we have

$$\int_0^b x^r f(x)\,dx = \int_0^{b^{1/k}} y^{kr} g(y)\,dy, \tag{34}$$

where $g(y) = ky^{k-1}f(y^k)$. Obviously $g(y)$ has, at least, contact of order $k-1$ at the start $y = 0$; further, if $f(x)$ has contact of order m at $x = b$, i.e. if $f(x) = O(x^{-m})$ at $x = b$, then $g(y) = O(y^{-(m-1)k-1})$. Hence there is high-order contact, of order $k-1$ and $(m-1)k+1$, respectively, at both ends of the range. The previous method is therefore applicable to $g(y)$ provided we can obtain its moments from those of $f(x)$. It is obvious from (34) that in order to obtain the ordinary moment of $g(y)$ we require to know the 'fractional' moments of $f(x)$, i.e. those corresponding to $r = j/k$ $(j = 0, 1, \ldots)$. If the moments of $f(x)$ are only known for integer r the fractional moments will have to be obtained by interpolation of the logarithmic moment function $\log \mu_r$ which will be more fully discussed in the next paper.

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
FISHER, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
FISHER, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
GEARY, R. C. (1947). *Biometrika*, **34**, 68.
GEARY, R. C. & WORLLEDGE, J. P. G. (1947). *Biometrika*, **34**, 98.
HARTLEY, H. O. (1940). *Biometrika*, **31**, 249.
KENDALL, M. G. (1938). *J.R. statist. Soc.* **101**, 592.
KENDALL, M. G. (1945). *The Advanced Theory of Statistics*, **1**. London: C. Griffin and Co.
MILNE THOMSON, L. M. (1933). *The Calculus of Finite Differences*. London: Macmillan.
NAIR, U. S. (1936). *Biometrika*, **30**, 274.
NAYER, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38.
NEYMAN, J. & PEARSON, E. S. (1931). *Bull. int. Acad. Cracovie*, A, p. 460.
SUKHATME, P. V. (1936). *Statist. Res. Mem.* **1**, 94.
Tables of the Incomplete Beta-Function, edited by Karl Pearson (1934). London: Biometrika Office.
WELCH, B. L. (1935). *Biometrika*, **27**, 145.
WELCH, B. L. (1936). *Statist. Res. Mem.* **1**, 1.

APPROXIMATION TO PERCENTAGE POINTS OF THE z-DISTRIBUTION

By A. H. CARTER, *King's College, Cambridge*

Tables have been published of the values of z for various percentage levels (20, 5, 1 and 0·1 %) for a range of given n_1, n_2 (Fisher & Yates, 1943, Table V). When n_1 or n_2 is outside the range of the tables, recourse must be had to approximate formulae (unless, of course, interpolation is sufficiently accurate) which will combine accuracy with facility of computation. One such formula, due to Fisher, with a modification suggested by Cochran (1940), is given at the foot of the above-mentioned tables. The purpose of this paper is to derive an alternative formula, no more difficult to compute, which will be shown to give consistently closer approximations to the true value of z for all except small n_1 or n_2.

Wishart (1947) has derived formulae for the exact cumulants of z, and also the well-known approximations to them when n_1, n_2 are large. The exact cumulants as far as κ_6 can be readily obtained arithmetically from tables of the Polygamma functions. Knowing the cumulants of the distribution, we may make use of the Cornish-Fisher normalization function method, based on Edgeworth's form of the Gram-Charlier type A series (Cornish & Fisher, 1937), to approximate to the percentage points. The method consists in writing z as an expansion in powers of a corresponding normal variate, ξ, the coefficients being functions of the cumulants of z, and assumes that κ_r is of order n^{1-r}, which is true for the z-distribution (Wishart, 1947, p. 172).

If z and ξ are expressed in standard measure (i.e. mean zero, standard deviation unity) we then derive

$$\frac{z-\mu_1'}{\sigma} = z' \sim \xi + \frac{\kappa_3}{\sigma^3}\frac{\xi^2-1}{6} + \frac{\kappa_4}{\sigma^4}\frac{\xi^3-3\xi}{24} - \frac{\kappa_3^2}{\sigma^6}\frac{2\xi^3-5\xi}{36},$$

correct to order n^{-1} for z'. This gives

Formula (a):
$$z \sim \mu_1' + \sigma\xi + \frac{\kappa_3}{\sigma^2}\frac{\xi^2-1}{6} + \frac{\kappa_4}{\sigma^3}\frac{\xi^3-3\xi}{24} - \frac{\kappa_3^2}{\sigma^5}\frac{2\xi^3-5\xi}{36}, \qquad (1)$$

correct to order n^{-1} (since $\sigma = O(n^{-\frac{1}{2}})$), where $\mu_1'(=k_1)$, $\sigma^2(=\kappa_2)$, κ_3, κ_4 are cumulants of the z-distribution. The ξ-coefficients may be readily computed: e.g. for the 5 % level, substitute $\xi = 1{\cdot}64485$. Table 2 gives the values, for the 20, 5, 1 and 0·1 % levels, of the coefficients required in applying the formula. The quantities μ_1', σ, κ_3, κ_4 depend of course on n_1 and n_2, and may be evaluated in any particular case, whence substitution in (1) gives the appropriate value of z. Since $|z_{1-P}(n_1, n_2)| = |z_P(n_2, n_1)|$, where z_P is the value of z corresponding to probability P, to find the percentage points for the 'negative tail', i.e. 80, 95, 99 and 99·9 %, we may simply interchange n_1 and n_2. This has the effect of changing the sign of the *odd* cumulants, so that in (1) we write $-\mu_1'$ and $-\kappa_3$ for μ_1' and κ_3.

Formula (a), being an approximation to order n^{-1}, may be expected to give reliable results when n_1 and n_2 are both large. For the 1 % point, for example, we find $z(6, 12) = 0{\cdot}7843$ (true value 0·7864), whereas $z(24, 60) = 0{\cdot}3744$ (true value 0·3746). Some further results for (6, 12), (6, 60), and (24, 60) are shown in Table 1 (a).

In practice, some labour is involved in applying formula (a), even if polygamma tables are available. The Fisher-Cochran formula, derived by the normalization function method, is a simple working approximation, valid for large n_1, n_2, in which the exact cumulants are replaced by their approximations in terms of inverse powers of n_1 and n_2.

Table 1. *Comparison of approximations to the percentage points of z*

Formula (*b*): Existing formula (Fisher-Cochran).
Formula (*c*): New formula.

Percentage level	$n_1, n_2 \rightarrow$	6, 12	6, 60	24, 60	20, 36	20, 100	36, 60	24, 24	36, 36
20	Formula (*b*)	0·2687	0·1901	0·1335	0·1577	0·1287	0·1212	0·1741	0·1415
	Formula (*c*)	0·2733	0·2020	0·1340	0·1580	0·1298	0·1213	0·1740	0·1415
	True *z*	0·2706	0·1965	0·1338	0·1579	0·1294	0·1213	0·1740	0·1415
5	Formula (*b*)	0·5507	0·3990	0·2650	0·3128	0·2573	0·2390	0·3426	0·2778
	Formula (*c*)	0·5501	0·4100	0·2654	0·3129	0·2586	0·2391	0·3426	0·2778
	True *z*	0·5487	0·4064	0·2654	0·3129	0·2583	0·2391	0·3425	0·2778
1	Formula (*b*)	0·7992	0·5646	0·3746	0·4441	0·3619	0·3385	0·4894	0·3955
	Formula (*c*)	0·7886	0·5698	0·3746	0·4435	0·3629	0·3384	0·4893	0·3955
	True *z*	0·7864	0·5687	0·3746	0·4435	0·3630	0·3384	0·4890	0·3954
0·1	Formula (*b*)	1·1074	0·7474	0·4963	0·5928	0·4755	0·4503	0·6602	0·5307
	Formula (*c*)	1·0693	0·7372	0·4954	0·5906	0·4756	0·4498	0·6595	0·5304
	True *z*	1·0628	0·7377	0·4955	0·5905	0·4760	0·4498	0·6589	0·5302

Percentage level	$n_1, n_2 \rightarrow$	12, 6	60, 6	60, 24	36, 20	100, 20	60, 36
20	Formula (*b*)	0·3509	0·3346	0·1566	0·1783	0·1656	0·1314
	Formula (*c*)	0·3506	0·3408	0·1569	0·1785	0·1665	0·1315
	True *z*	0·3510	0·3388	0·1568	0·1784	0·1661	0·1315
5	Formula (*b*)	0·6884	0·6435	0·3047	0·3483	0·3208	0·2566
	Formula (*c*)	0·7001	0·6706	0·3060	0·3493	0·3236	0·2570
	True *z*	0·6931	0·6596	0·3055	0·3488	0·3227	0·2568
1	Formula (*b*)	1·0120	0·9444	0·4368	0·4995	0·4615	0·3661
	Formula (*c*)	1·0370	0·9956	0·4391	0·5016	0·4662	0·3667
	True *z*	1·0218	0·9770	0·4385	0·5009	0·4666	0·3666
0·1	Formula (*b*)	1·4352	1·3340	0·5930	0·6789	0·6303	0·4932
	Formula (*c*)	1·4681	1·4155	0·5965	0·6820	0·6375	0·4942
	True *z*	1·4449	1·3929	0·5962	0·6814	0·6371	0·4940

Table 1 (*a*). *Some values of z from formula* (*u*) (*exact cumulant formula*)

(For corresponding true values, see Table 1)

% \ n_1, n_2	6, 12	6, 60	24, 60	12, 6	60, 6	60, 24
20	0·2699	0·1998	0·1338	0·3499	0·3335	0·1567
5	0·5457	0·4022	0·2652	0·6958	0·6627	0·3057
1	0·7843	0·5640	0·3744	1·0295	0·9854	0·4388
0·1	1·0684	0·7433	0·4956	1·4592	1·4026	0·5966

The cumulant function of z is

$$K(z) = \tfrac{1}{2}it\log\frac{n_2}{n_1} + \log\Gamma\left(\frac{n_1+it}{2}\right) + \log\Gamma\left(\frac{n_2-it}{2}\right) - \log\Gamma\left(\frac{n_1}{2}\right) - \log\Gamma\left(\frac{n_2}{2}\right),$$

and the cumulants are obtainable by differentiating this successively with respect to (it), at each stage putting $t = 0$.

Since $$\left[\frac{d^r}{d(it)^r}\log\Gamma\left(\frac{n \pm it}{2}\right)\right]_{t=0} = (\pm 1)^r \frac{d^r}{dn^r}\log\Gamma\left(\frac{n}{2}\right),$$

and $\log\Gamma(\frac{1}{2}n)$ can be expanded by Stirling's theorem in inverse powers of n, the cumulants may also readily be expressed in inverse powers of n_1, n_2; and, when n_1, n_2 are reasonably large, the first few terms only in the expansions will give sufficiently close approximations to them. In fact, writing

$$\frac{1}{n_1}+\frac{1}{n_2} = s, \quad \frac{1}{n_1}-\frac{1}{n_2} = d,$$

it has been shown that

$$\left.\begin{aligned} \kappa_1 &= \mu_1' = -\tfrac{1}{2}d - 6sd + O(n^{-4}), \\ \kappa_2 &= \sigma^2 = \tfrac{1}{2}s + \tfrac{1}{4}(s^2+d^2) + O(n^{-3}), \\ \kappa_3 &= -\tfrac{1}{2}sd + O(n^{-3}), \\ \kappa_4 &= \tfrac{1}{4}s(s^2+3d^2) + O(n^{-4}), \\ \kappa_r &= O(n^{1-r}) \quad (r>1). \end{aligned}\right\} \tag{A}$$

Formula (a) will now have an extra term, since we take as our 'working variance' of z not its exact value, but its approximation to order n^{-1} from (A), i.e. $\frac{1}{2}s$. In the notation of Kendall (1945)

$$l_2 = \kappa_2/\tfrac{1}{2}s - 1 \sim \tfrac{1}{2}(s+d^2/s).$$

We then obtain

$$z - \mu_1' \sim \sqrt{\left(\frac{s}{2}\right)}\xi - \frac{d}{6}(\xi^2-1) + \sqrt{\frac{s}{2}}\left\{\frac{s}{24}(\xi^3+3\xi) + \frac{d^2}{72s}(\xi^3+11\xi)\right\} \tag{2}$$

$$= \frac{\xi}{\sqrt{2/s}}\left\{1 + \frac{1}{2/s}\frac{\xi^2+3}{12}\right\} - \frac{d}{6}(\xi^2-1) + \frac{d^2}{144}\sqrt{\frac{2}{s}}(\xi^3+11\xi)$$

or

$$z - \mu_1' \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2-1), \tag{3}$$

where

$$h = \frac{2}{s}, \quad \lambda = \frac{\xi^2+3}{6},$$

provided $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$ may be neglected (which will be the case for small d). Inserting the approximation to μ_1' from (A), i.e. $-\frac{1}{2}d$,

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2+2), \tag{3a}$$

the Fisher-Cochran formula, which has, in fact, been found to give a fairly close approximation to the true z for n_1, n_2 both reasonably large. It may be noted that if n_1, n_2 are not very large, an improvement will be effected by including the second term in the estimate of the mean (κ_1), i.e. from (A) by adding $-\frac{1}{2}sd$.

For $(n_1, n_2) = (6, 12)$ this correction is $-0{\cdot}00347$, and for $(24, 60)$, the correction is $-0{\cdot}00024$. Inserting this improved approximation to μ_1' in (3) we have

Formula (b):
$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2+2+s). \tag{3b}$$

As pointed out by Wishart (1947, p. 179), an approximation to the value of any $\kappa_r (r > 1)$ obtained by considering its leading term only, will be improved by writing $1/(n_1 - 1)$ and $1/(n_2 - 1)$ in place of $1/n_1$ and $1/n_2$. For, by Stirling's expansion of a factorial,

$$\log \Gamma\left(\frac{n}{2}\right) = \frac{n-1}{2}\log\frac{n}{2} - \frac{n}{2} + \tfrac{1}{2}\log 2\pi + \frac{1}{6n} + O(n^{-3}).$$

$$\frac{d}{dn}\log \Gamma\left(\frac{n}{2}\right) = \tfrac{1}{2}\log n - \frac{1}{2n} - \frac{1}{6n^2} + O(n^{-4}),$$

$$\frac{d^2}{dn^2}\log \Gamma\left(\frac{n}{2}\right) = \frac{1}{2n} + \frac{1}{2n^2} + \frac{1}{3n^3} + O(n^{-5})$$

$$= \frac{1}{2(n-1)} + O(n^{-3}),$$

$$\frac{d^3}{dn^3}\log \Gamma\left(\frac{n}{2}\right) = -\frac{1}{2(n-1)^2} + O(n^{-4}),$$

and so on.

Table 2. *ξ-coefficients required in applying formula (a)*

	20%	5%	1%	0·1%
ξ	0·84162	1·64485	2·32635	3·09023
$\frac{1}{6}(\xi^2-1)$	−0·04861	0·28426	0·73532	1·42492
$\frac{1}{24}(\xi^3-3\xi)$	−0·08036	−0·02018	0·23379	0·84332
$\frac{1}{36}(2\xi^3-5\xi)$	−0·08377	0·01878	0·37634	1·21026

Thus writing $$s' = \frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}, \quad d' = \frac{1}{n_1 - 1} - \frac{1}{n_2 - 1},$$

we might expect a better approximation to z to be obtained, corresponding to that of Fisher and Cochran, if we use s' and d' instead of s and d.

Corresponding to equations (A) we have

$$\left.\begin{aligned} \kappa_2 &= \tfrac{1}{2}s' - \tfrac{1}{24}s'(s'^2 + 3d'^2) + O(n^{-4}), \\ \kappa_3 &= -\tfrac{1}{2}s'd' + O(n^{-4}), \\ \kappa_4 &= \tfrac{1}{4}s'(s'^2 + 3d'^2) + O(n^{-5}), \\ \kappa_r &= O(n^{1-r}) \quad (r > 1). \end{aligned}\right\} \tag{B}$$

For the mean, however, $$\mu_1' = \kappa_1 = -\tfrac{1}{2}d - \tfrac{1}{8}sd + O(n^{-4}), \tag{4}$$

$$= -\tfrac{1}{2}d' + \tfrac{1}{8}s'd' + O(n^{-3}), \tag{4a}$$

If n^{-3} is *not* negligible (relative to the degree of accuracy desired) μ_1' should therefore be left in the form $-\frac{1}{2}d - \frac{1}{8}sd$.

Proceeding as before, we obtain

$$z - \mu_1' \sim \sqrt{\left(\frac{s'}{2}\right)}\,\xi - \frac{d'}{6}(\xi^2 - 1) + s'\sqrt{\left(\frac{s'}{2}\right)}\frac{\xi^3 - 3\xi}{24} + d'^2\sqrt{\left(\frac{2}{s'}\right)}\frac{\xi^3 - 7\xi}{144}, \tag{5}$$

whence $$z - \mu_1' \sim \frac{\xi}{\sqrt{(h' - \lambda')}} - \frac{d'}{6}(\xi^2 - 1), \tag{6}$$

where $$h' = 2/s', \quad \lambda' = \tfrac{1}{6}(\xi^2 - 3).$$

Since this is based in the first place on more accurate approximations to the cumulants κ_3 and κ_4, and since the term omitted from (5) in deriving (6) (i.e. $\frac{1}{144}d'^2\sqrt{(2/s')}(\xi^3-7\xi)$, is evidently numerically less than the corresponding term omitted in obtaining the Fisher-Cochran formula (i.e. $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$, formula (6) might be expected to give an improved approximation to z. In fact, however, it does not, and the reason is not far to seek.

Consider the expansion of $\xi/\sqrt{(h-\lambda)}$ in both cases:

$$\frac{\xi}{\sqrt{(h-\lambda)}} = \frac{\xi}{\sqrt{h}}\left(1+\frac{\lambda}{2h}+\frac{3\lambda^2}{4h^2}+\ldots\right),$$

where the terms are decreasing in magnitude (since $1/h = \frac{1}{2}s = O(n^{-1})$). Hence the error in neglecting all terms after the second will be approximately of the order of the third term. Now in the Fisher-Cochran approximation this term, $+\frac{3\lambda^2\xi}{4h^2\sqrt{h}}$, has the same sign as the omitted term $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$ (both being of the same sign as ξ), so that the extra terms included will tend to compensate for the term omitted. In obtaining (6) from (5), on the other hand, the term omitted, $\frac{1}{144}d'^2\sqrt{(2/s')}(\xi^3-7\xi)$, will be of *opposite* sign to ξ when $|\xi|<\sqrt{7}$, corresponding to a probability of about 0·004: so that for most percentage levels encountered in practice, the error in (5) is increased in (6).

A better formula is obtained from (5) as

$$z-\mu_1' \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \frac{d'}{6}(\xi^2-1), \tag{7}$$

where h' and λ' are as in (6).

Expanding $\xi\sqrt{(h'+\lambda')}/h'$, it is found that the third term is now of *opposite* sign to ξ, and hence the extra terms contained in the expansion will tend to compensate for the term omitted. Since s' and d' require to be calculated in applying this formula, it is desirable to write μ_1' in the form (4*a*) (provided we can neglect quantities of order n^{-3}). This gives

Formula (*c*): $$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \frac{d'}{6}(\xi^2+2-2s'). \tag{7a}$$

Collecting the results, we have the three approximate formulae:

Formula (*a*) (exact cumulant method):

$$z \sim \kappa_1+\sigma\xi+\frac{\kappa_2}{\sigma^2}\frac{\xi^2-1}{6}+\frac{\kappa_4}{\sigma^3}\frac{\xi^3-3\xi}{24}-\frac{\kappa_3^2}{\sigma^5}\frac{2\xi^3-5\xi}{36}.$$

Formula (*b*) (Fisher-Cochran formula):

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2+2+s),$$

where $$s=\frac{1}{n_1}+\frac{1}{n_2},\quad d=\frac{1}{n_1}-\frac{1}{n_2},\quad h=\frac{2}{s},\quad \lambda=\frac{\xi^2+3}{6}.$$

Formula (*c*) (new formula): $$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \frac{d'}{6}(\xi^2+2-2s'),$$

where $$s'=\frac{1}{n_1-1}+\frac{1}{n_2-1},\quad d'=\frac{1}{n_1-1}-\frac{1}{n_2-1},\quad h'=\frac{2}{s'},\quad \lambda'=\frac{\xi^2-3}{6}.$$

When n_1, n_2 are large (and not too different), $\frac{1}{6}sd$ is negligible and formula (b) becomes the formula more generally quoted

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)\frac{\xi^2+2}{6}$$

which may be written

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)(\lambda - \tfrac{1}{6}).$$

Similarly, for sufficiently large n_1, n_2, $\frac{1}{6}s'd'$ may be neglected and formula (c) becomes

$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)\frac{\xi^2+2}{6},$$

or

$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)(\lambda' + \tfrac{5}{6}).$$

It is to be noted, however, that since $\frac{1}{6}s'd'$ is approximately twice $\frac{1}{6}sd$, more care must be exercised in deciding to neglect it. For example, when (n_1, n_2) is $(20, 100)$, $\frac{1}{6}s'd' = 0{\cdot}0009$, and for $(24, 60)$, its value is $0{\cdot}0005$.

For purposes of comparison, values of z have been computed from formulae (b) and (c), for the four common percentage levels, over a fairly wide range of n_1, n_2. They are shown in Table 1, together with the corresponding true values of z. The latter were obtained where possible from the tables of Fisher and Yates: elsewhere by inverse interpolation in *Tables of the Incomplete Beta-Function* followed by a logarithmic transformation. Such values are in error by not more than $0{\cdot}0001$. It will be seen that neither formula yields very accurate results when n_1 or n_2 is as small as 6, though even here the new formula is rather better with the single exception of $n_1 = 12$, $n_2 = 6$. In actual practice, however, we are concerned with large values of n_1, n_2, beyond the range of the published tables. Considering only those cases where n_1 and n_2 are both greater than 20, it is seen that formula (c) gives a consistently closer approximation than does formula (b) for both the positive and the negative tails, and for all the percentage levels investigated, though its relative gain in accuracy is greatest at the 1 and $0{\cdot}1$ % levels. It may be noted, in fact, that in no case considered having n_1 and n_2 greater than 20, is the error more than 9 in the fourth decimal place, i.e. it appears that for all except small n_1, n_2, this formula will give an approximation to z correct to within $0{\cdot}001$.

In conclusion, therefore, it is recommended that formula (c) be adopted for general use, since it is no more difficult to compute, and is more accurate, than the existing formula. Dropping the dashes we have the formula

$$z \sim \frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)\left(\lambda + \frac{5}{6} - \frac{s}{3}\right),$$

where

$$s = \frac{1}{n_1-1} + \frac{1}{n_2-1}, \quad h = \frac{2}{s}, \quad \lambda = \frac{\xi^2-3}{6}$$

or, if $\frac{1}{3}\left\{\frac{1}{(n_1-1)^2} - \frac{1}{(n_2-1)^2}\right\}$ may be neglected,

$$z \sim \frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)(\lambda + \tfrac{5}{6}).$$

The values of ξ and λ for the four percentage levels are:

	20%	5%	1%	0·1%
ξ	0·8416	1·6449	2·3263	3·0902
λ	−0·3819	0·0491	0·4020	1·0916

My thanks are due to Dr J. Wishart, whose suggestion was the basis of this paper.

REFERENCES

Cochran, W. G. (1940). Note on an approximative formula for significance levels of z. *Ann. Math. Statist.* **11**, 93.

Cornish, E. A. & Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **4**, 307.

Fisher, R. A. & Yates, F. (1943). *Statistical Tables*, 2nd ed. Oliver and Boyd.

Kendall, M. G. (1945). *Advanced Theory of Statistics*, **1**, 156, 2nd ed. C. Griffin and Co.

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions. *Biometrika*, **34**, 170.

MISCELLANEA

Note on the cumulants of Fisher's z-distribution

By LEO A. AROIAN, *Hunter College*

In a recent article Dr J. Wishart (1947) stated: 'Explicit expressions for the exact cumulants of Fisher's z-distribution do not appear ever to have been published.' Fisher's z-distribution and the related Snedecor's F-distribution formed a part of my doctor's thesis and rather full results concerning the cumulants of the z-distribution and other properties of the distribution were published in the *Annals of Mathematical Statistics* (Aroian, 1941) some time ago.* I should like to take this opportunity of adding certain comments on the Gram-Charlier Type A approximation to the z-distribution and the type III approximation to the F-distribution.

To obtain the cumulants of the z-distribution I expanded the moment generating function $M_z(\theta)$ in powers of θ and found $\lambda_{k:z}$, the kth semi-variant (or cumulant) of z as the coefficient of $\theta^k/k!$. The exact results correspond with Wishart's formulae (9) to (15), although given in a different form, and need not be repeated here. In addition, asymptotic formulae for $\lambda_{k:z}$, n_1 and n_2 large, were derived by means of the Euler-Maclaurin sum formula. Furthermore, another type of formula could have been given for n_1 small but n_2 large, merely by expanding that part of $\lambda_{k:z}$ in which n_2 occurs by the Euler-Maclaurin sum formula. The special cases for the logarithmic χ^2, the logarithmic t, and the logarithmic normal probability functions follow by substituting the proper limiting values of n_1 and n_2.

In my previous paper I was overcautious concerning the type A approximation to the z-distribution. Actually the method is fairly accurate although tedious. Taking

$$F(t) = \phi(t) + A_3\phi^{iii}(t) + A_4\phi^{iv}(t), \quad \int_{t_0}^{\infty} F(t)\,dt = \eta,$$

we have

$$\int_{t_0}^{\infty} F(t)\,dt = \int_{t_0}^{\infty} \phi(t)\,dt + \phi(t)\{-A_3(t_0^2-1) + A_4(t_0^3-3t_0)\},$$

where η is usually 0·10, 0·05, 0·025, 0·01, etc. As an example take $n_1 = 24$, $n_2 = 60$; then

$$\lambda_{1:z} = -0{\cdot}0127429, \quad \sigma_z = 0{\cdot}173779, \quad \lambda_{3:z} = -0{\cdot}0007998, \quad \lambda_{4:z} = 0{\cdot}0000867,$$

$$A_3 = \frac{-\lambda_{3:z}}{3!\sigma_z^3} = 0{\cdot}025345, \quad A_4 = \frac{\lambda_{4:z}}{4!\sigma_z^4} = 0{\cdot}00396.$$

t_0 for $\eta = 0{\cdot}05$ is 1·60094, $z_{0{\cdot}05} = 0{\cdot}26547$ against the accurate value of 0·26534. For the 1 % point $t_0 = 2{\cdot}2338$, $z_{0{\cdot}01} = 0{\cdot}3754$ against the accurate value of 0·3746. When $n_1 = n_2 = 24$, $z_{0{\cdot}05} = 0{\cdot}3423$ against the accurate value of 0·3425.

The type III approximation to the F-distribution is of some interest since for n_1 moderate and n_2 large, n_1F tends to be distributed as χ^2 with n_1 degrees of freedom. Since

$$\text{Mean } F = \bar{F} = \frac{n_2}{n_2-2}; \quad \sigma_F = \frac{n_2}{n_2-4}\sqrt{\frac{2(n_1+n_2-2)}{n_1(n_2-4)}},$$

$$\alpha_{3:F} = \frac{4(2n_1+n_2-2)}{n_1(n_2-6)}\sqrt{\frac{n_1(n_2-4)}{2(n_1+n_2-2)}}, \quad \alpha_{3:F} = \sqrt{(\beta_{1:F})},$$

we find the 5, 1 or 0·1 % points for F by using

$$F_{0{\cdot}05} = \bar{F} + \sigma_F(1{\cdot}64485 + 0{\cdot}28392\alpha_{3:F} - 0{\cdot}04902\alpha_{3:F}^2),$$

$$F_{0{\cdot}01} = \bar{F} + \sigma_F(2{\cdot}32635 + 0{\cdot}73330\alpha_{3:F} - 0{\cdot}024957\alpha_{3:F}^2),$$

$$F_{0{\cdot}001} = \bar{F} + \sigma_F(3{\cdot}0903 + 1{\cdot}4190\alpha_{3:F} + 0{\cdot}05667\alpha_{3:F}^2).$$

* [Both Dr Wishart, as author, and myself as editor regret that owing to wartime preoccupation the publication of Dr Aroian's 1941 paper was overlooked. E.S.P.]

These formulae for the levels of significance of the χ^2 distribution are from a previous paper (Aroian, 1943). For $n_1 = 24$, $n_2 = 60$, $F_{0{\cdot}05}$ by this approximation is 1·709 compared with the accurate value of 1·700. For $n_1 = 24$, $n_2 = 100$, $F_{0{\cdot}05}$ by this approximation is 1·631 against the accurate value of 1·627. For $n_1 = n_2 = 100$, $F_{0{\cdot}05}$ by this approximation is 1·394 as compared with the accurate value of 1·392. While these results are not too poor, they are not so accurate as the well-known formulae of Cochran-Fisher or of E. Paulson (1942) which, for large values of n_1 and n_2, generally give 4 significant figures.

REFERENCES

Aroian, L. A. (1941). *Ann. Math. Statist.* **12**, 429.
Aroian, L. A. (1943). *Ann. Math. Statist.* **14**, 93.
Paulson, E. (1942). *Ann. Math. Statist.* **13**, 233.
Wishart, J. (1947). *Biometrika*, **34**, 170.

A note on the mean deviation from the median

By K. R. NAIR

For samples drawn from a normal universe, Godwin (1945) obtained the sampling distribution of the mean deviation when the individual deviations are measured from the sample *mean*. It is well known that the mean deviation is least when it is measured from the sample *median*.* Let us refer to them as 'mean deviation from mean' and 'mean deviation from median' respectively, and use the letters m and m' to denote their sample estimates.

The exact sampling distribution of m being now known and its probability integral tabulated, the question may well be asked what the distribution of m' is. Since $m' \leqslant m$, their expectations have the same relationship

$$E(m') \leqslant E(m). \tag{1}$$

For samples of n from a normal population with standard deviation σ, $E(m') = f'_n\sigma$ and $E(m) = f_n\sigma$, where $f'_n \leqslant f_n$. For getting unbiased estimates of σ we should divide m' by f'_n and m by f_n. What we are now interested to know is which of the two estimates has a smaller standard error. In the case of m, it has been shown by Helmert (1876) and Fisher (1920) that

$$f_n = \sqrt{\frac{2(n-1)}{n\pi}}, \tag{2}$$

and

$$\text{S.E. of } \left(\frac{m}{f_n}\right) = \frac{\sigma}{\sqrt{n}}\sqrt{\left[\left\{\frac{\pi}{2} + \sqrt{[n(n-2)]} - n + \sin^{-1}\frac{1}{n-1}\right\}\right]}. \tag{3}$$

In the case of m', we neither know f'_n nor the standard error of (m'/f'_n) for samples of size n.

It is obvious that when n is very large, the mean and median will differ very little from one another and hence $m' \to m$ and $f'_n \to f_n$. It is interesting to note that, at the other end of the scale, namely, when $n = 2$, m and m' are identical, and equal to one-half the sample range.

To discover any real difference that may exist between the standard errors of (m/f_n) and (m'/f'_n), which is the same as determining the difference between the coefficients of variation of m and m', we must consider samples of size greater than 2.

(i) Let us take $n = 3$, and let x_1, x_2, x_3 be the observed values arranged in order of ascending magnitude. We at once find that

$$m' = \tfrac{1}{2}(x_3 - x_1). \tag{4}$$

The distribution of m' for samples of 3 is therefore derivable from that of the range. The probability integral of the range has been tabulated by Pearson & Hartley (1942) for $n = 2$ to 20. For our purpose it is necessary only to know the values of the mean range $(\overline{w})$ and the standard error of the range (σ_w).

* When n, the sample size, is an odd number, the sample median is by definition the value of the $\frac{1}{2}(n+1)$th ranked observation. When n is even, the sample median is conventionally taken as the mean of the $\frac{1}{2}n$th and $\frac{1}{2}(n+2)$th ranked values. The mean deviation from the median will have the same magnitude whatever value, between the $\frac{1}{2}n$th and $\frac{1}{2}(n+2)$th ranked values, the median takes, when n is even. No complication is therefore introduced by accepting the conventional definition of the median for even-sized samples.

for samples of 3. This can be calculated, correct to six decimal places, from certain numerical values given by Pearson (1926). Using his figures,

$$\overline{w} = 1{\cdot}692568 \times \sigma, \tag{5}$$

$$\sigma_w = 0{\cdot}888368 \times \sigma. \tag{6}$$

The value of f'_n for sample of 3 is, therefore,

$$f'_3 = \tfrac{1}{2} \times 1{\cdot}692568 = 0{\cdot}56419,$$

and the standard error of m'/f'_3 is $$\frac{0{\cdot}888368}{1{\cdot}692568}\sigma = 0{\cdot}52486\sigma, \tag{7}$$

correct to five decimal places.

The corresponding values for f_3 and standard error of (m/f_3) are obtained by putting $n = 3$ in equations (2) and (3) and are

$$f_3 = \sqrt{\frac{4}{3\pi}} = 0{\cdot}65147, \tag{8}$$

and $$\text{S.E. of } (m/f_3) = \frac{\sigma}{\sqrt{3}}\sqrt{\left(\frac{2\pi}{3} + \sqrt{3} - 3\right)} = 0{\cdot}52486\sigma, \tag{9}$$

correct to five decimal places.

Although (9) can be evaluated to any number of decimal places, we are not in a position to bring (7) to a higher order of accuracy than five decimal places. It is very unlikely that (7) and (9) are absolutely identical, but we may safely conclude that they are practically the same.

(ii) We next come to samples of 4. If x_1, x_2, x_3, x_4 be the observations arranged in order of ascending magnitude, the mean deviation from median is given by

$$m' = \tfrac{1}{2}(x_4 + x_3 - x_2 - x_1). \tag{10}$$

The distribution of m' follows immediately from 'some order statistic distributions for samples of size 4' obtained by Walsh (1946) and is as follows:

$$p(m')\,dm' = \frac{12}{[\sqrt{(2\pi)}]^3} e^{-2m'^2}\left(\int_0^{4m'} e^{-\frac{1}{2}y^2}dy\right)^2 dm'. \tag{11}$$

The probability integral of m' is given by

$$P(m') = \int_0^{m'} p(m')\,dm' = \left(\int_0^{4m'} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y^2}dy\right)^3 = \left(\frac{2}{\sqrt{(2\pi)}}\int_0^{2m'} e^{-\frac{1}{2}x^2}dx\right)^3. \tag{12}$$

The values of $P(m')$ given by (12) can easily be evaluated using the normal probability integral table and are given in cols. (3) and (6) of the table below, alongside corresponding values (given in cols. (2) and (5)) for the probability integral of the mean deviation (m) from the mean, for samples of 4, copied from Godwin's (1945) tables.

Table giving the probability integral of the mean deviation from (a) mean and (b) median for samples of four observations from a normal universe ($\sigma = 1$)

m (or m')	$P(m)$	$P(m')$	m (or m')	$P(m)$	$P(m')$
0·0	0·00000	0·00000	1·3	0·96758	0·97229
0·1	0·00333	0·00398	1·4	0·98229	0·98475
0·2	0·02534	0·03003	1·5	0·99073	0·99192
0·3	0·07879	0·09204	1·6	0·99534	0·99588
0·4	0·16693	0·19139	1·7	0·99775	0·99798
0·5	0·28345	0·31818	1·8	0·99895	0·99905
0·6	0·41552	0·45629	1·9	0·99953	0·99957
0·7	0·54836	0·58951	2·0	0·99980	0·99981
0·8	0·66934	0·70592	2·1	0·99992	0·99992
0·9	0·77040	0·79954	2·2	0·99997	0·99997
1·0	0·84860	0·86962	2·3	0·99999	0·99999
1·1	0·90502	0·91888	2·4	1·00000	1·00000
1·2	0·94321	0·95162			

We note that although m and m' have an infinite range from 0 to ∞, their probability integrals rapidly approach unity, this value being reached to five decimal place accuracy when $m(m') = 2{\cdot}4\sigma$. We can approximately work out the moments of the two distributions from the table above. The values of the mean and the standard deviation (applying Sheppard's correction for grouping) of m and m' so obtained are given below:

$$\left.\begin{array}{lll} \text{Mean:} & \overline{m} = 0{\cdot}690986\sigma, & \overline{m}' = 0{\cdot}663187\sigma, \\ \text{Standard deviation:} & \sigma_m = 0{\cdot}297015\sigma, & \sigma_{m'} = 0{\cdot}292979\sigma, \\ \text{Coefficient of variation:} & \sigma_m/\overline{m} = 0{\cdot}429842, & \sigma_{m'}/\overline{m}' = 0{\cdot}441775. \end{array}\right\} \tag{13}$$

The values of $\overline{m}$ and $\sigma_m/\overline{m}$ obtained from the exact formulae (2) and (3) are

$$\left.\begin{aligned} \overline{m} &= \sigma\sqrt{\left(\frac{3}{2\pi}\right)} = 0{\cdot}690988\sigma, \\ \sigma_m/\overline{m} &= \tfrac{1}{2}\sqrt{(\tfrac{1}{2}\pi + \sin^{-1}\tfrac{1}{3} + 2\sqrt{2} - 4)} = 0{\cdot}429842, \end{aligned}\right\} \tag{14}$$

showing close agreement with the values given in (13) for the mean and coefficient of variation of m. We may therefore consider the mean and coefficient of variation of m', approximately evaluated in (13), to be of sufficient accuracy to warrant the conclusion that for samples of size 4, the mean deviation from the mean leads to a more 'efficient' estimate of the population standard deviation than the mean deviation from the median. As the distribution of the latter is not known for $n > 4$, we are not in a position to say whether this conclusion holds good, in general, for all values of n.

In conclusion, it seems worth making the following point:

(*a*) if expressions for the expectation and variance of m' were available and tables of its probability integral worked out,

(*b*) if the efficiency of the m' estimate compared to the m estimate for $n > 4$ was not appreciably worse than for the case $n = 4$,

there would be strong practical grounds for using m' rather than m in view of greater simplicity in calculation. In both cases we must first arrange the observations in order of magnitude. Then if $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, m' may be calculated from the formula

$$m' = \frac{1}{t}\{(x_{n-t+1} + x_{n-t+2} + \ldots + x_n) - (x_1 + x_2 + \ldots + x_t)\}, \tag{15}$$

where $t = \frac{1}{2}n$ or $\frac{1}{2}(n-1)$ according as n is even or odd.

For m, however, we must also calculate the arithmetic mean $\bar{x}$ and look for x_k and x_{k+1} between which $\bar{x}$ lies. Then m can be obtained from one of the three formulae

$$\left.\begin{aligned} \frac{nm}{2k} &= \bar{x} - \frac{x_1 + x_2 + \ldots + x_k}{k}, \\ \frac{nm}{2(n-k)} &= \frac{x_{k+1} + \ldots + x_n}{n-k} - \bar{x}, \\ \frac{n^2m}{2k(n-k)} &= \frac{x_{k+1} + \ldots + x_n}{n-k} - \frac{x_1 + \ldots + x_k}{k}. \end{aligned}\right\} \tag{16}$$

This certainly involves a rather longer process.

It is interesting to note that m' becomes a special case of the measure of dispersion based on difference between the sums of the first and the last r observations (in order of magnitude) suggested by Jones (1946), the range, becoming another special case of the same measure, when $r = 1$.

REFERENCES

Fisher, R. A. (1920). *Mon. Not. R. Astr. Soc.* **80**, 758.
Godwin, H. J. (1945). *Biometrika*, **33**, 254.
Helmert, W. (1876). *Astr. Nachr.* **88**, no. 2096.
Jones, A. E. (1946). *Biometrika*, **33**, 274.
Pearson, E. S. (1926). *Biometrika*, **18**, 173.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 301.
Walsh, J. E. (1946). *Ann. Math. Statist.* **17**, 246.

On the method of paired comparisons

By P. A. P. MORAN, *Institute of Statistics, Oxford University*

M. G. Kendall & B. Babington Smith (1940) have discussed the 'method of paired comparisons' for investigating preferences. Suppose we are given n objects $A, \ldots, K$, and an observer is asked to choose between every pair. If A is preferred to B we write $A \to B$. If the observer is not completely consistent, either because of his own inefficiency or because the objects are not really capable of being ranked in respect of the quality under consideration, he may make preferences of the type $A \to B \to C \to A$, and we call this an inconsistent or circular triad. Write d for the number of circular triads in a given experiment. Then Kendall & Babington Smith show that

$$\zeta = 1 - \frac{24d}{n^3 - n} \quad (n \text{ odd})$$

$$= 1 - \frac{24d}{n^3 - 4n} \quad (n \text{ even})$$

may be regarded as a 'coefficient of consistence' and lies between 0 and 1, being capable of attaining both these limits.

Now suppose that each comparison is made at random so that there are equal chances that $A \to B$ and $B \to A$. The distribution of d is then of interest. They calculate this distribution exactly for $n = 2, \ldots, 7$ and conjecture that its moments are given by

$$\mu_1' = \frac{1}{4}\binom{n}{3},$$

$$\mu_2 = \frac{3}{16}\binom{n}{3},$$

$$\mu_3 = -\frac{3}{32}\binom{n}{3}(n-4),$$

$$\mu_4 = \frac{3}{256}\binom{n}{3}\left\{9\binom{n-3}{3} + 39\binom{n-3}{2} + 9\binom{n-3}{1} + 7\right\},$$

these being polynomials in n which agree with their numerical calculations for $n = 2, \ldots, 7$. They also conjecture that the distribution tends to normality when n increases. In the present note we prove these statements.

Let the objects be numbered from 1 to n. Write $P_{ijk} = 1$ if the triad (i, j, k) is circular, and $P_{ijk} = 0$ if it is not. Then $d = \Sigma P_{ijk}$, the sum being taken over all such triads. Now by enumerating the various cases we see that $E(P_{ijk}) = \frac{1}{4}$ and so $\mu_1'(d) = E(\Sigma P_{ijk}) = \frac{1}{4}\binom{n}{3}$. Now consider $\mu_2'(d) = E[(\Sigma P_{ijk})^2]$. Consider the types of terms which results when we expand this. In the first place we have $\binom{n}{3}$ terms typified by P_{123}^2, and these contribute $\frac{1}{4}\binom{n}{3}$ to $\mu_2'(d)$. Similarly, we have terms typified by $P_{123}P_{145}$, $P_{123}P_{124}$ and $P_{123}P_{456}$, and the number of these are respectively $\frac{3}{2}\binom{n}{3}(n-3)(n-4)$, $3\binom{n}{3}(n-3)$ and $\binom{n}{3}\binom{n-3}{3}$, whilst their expectations are each $\frac{1}{16}$. It follows that

$$\mu_2'(d) = \frac{1}{16}\binom{n}{3}\left\{\binom{n}{3} + 3\right\},$$

and so

$$\mu_2(d) = \frac{3}{16}\binom{n}{3}.$$

The calculation of $\mu_3'(d)$ is a good deal more complicated. $\mu_3'(d) = E((\Sigma P_{ijk})^3)$, and on expanding we get 16 types of terms, typified by

$$P_{123}^3, \quad P_{123}^2 P_{145}, \quad P_{123}^2 P_{124}, \quad P_{123}^2 P_{456},$$
$$P_{123} P_{145} P_{167}, \quad P_{123} P_{124} P_{135}, \quad P_{123} P_{145} P_{456}, \quad P_{123} P_{145} P_{146},$$
$$P_{123} P_{124} P_{145}, \quad P_{123} P_{234} P_{134}, \quad P_{123} P_{345} P_{567}, \quad P_{123} P_{345} P_{346},$$
$$P_{123} P_{145} P_{678}, \quad P_{123} P_{124} P_{567}, \quad P_{123} P_{456} P_{789}, \quad P_{123} P_{345} P_{345}.$$

After some calculation we find the sum of the contributions of these to be

$$\frac{1}{2304}\binom{n}{3}\{n^6 - 6n^5 + 13n^4 + 42n^3 - 158n^2 - 108n + 864\}.$$

Reducing to the mean we get $\mu_3(d) = -\frac{3}{32}\binom{n}{3}(n-4)$.

The calculation of $\mu_4'(d)$ is a great deal more complicated, there being 85 terms which are not zero; we finally obtain, after lengthy calculations,

$$\mu_4' = \frac{1}{55296}\binom{n}{3}\{n^9 - 9n^8 + 33n^7 + 45n^6 - 582n^5 + 504n^4 + 5732n^3 - 10692n^2 - 30024n + 80352\},$$

and so

$$\mu_4 = \frac{1}{55296}\binom{n}{3}\{972n^3 + 972n^2 - 26936n + 80352\},$$

which reduces to the conjectured result.

We now prove that the distribution tends to normality. To do this, it is sufficient (Kendall, 1943, p. 110) to prove that

$$\frac{\mu_{2m}}{\mu_2^m} \to \frac{(2m)!}{2^m m!}, \quad \frac{\mu_{2m+1}}{\mu_2^{\frac{1}{2}(2m+1)}} \to 0, \quad \text{for} \quad m = 1, 2, \ldots.$$

Consider the second of these first. Write $Q_{ijk} = P_{ijk} - \frac{1}{4}$. Then

$$\mu_{2m+1}(d) = E[(\Sigma Q_{ijk})^{2m+1}].$$

It is clear that for any given m we could calculate $\mu_{2m+1}(d)$ given sufficient labour, by expanding this and considering the expectation of each type of term and calculating the number of times it occurs, which will be a polynomial in n. Now consider the various types of terms in the expansion of $(\Sigma Q_{ijk})^{2m+1}$. We classify these terms according to whether the Q's have common suffixes. Let $Q_{ijk} Q_{lmn} \ldots Q_{pqr}$ be a typical product in the expansion. If this can be separated into p groups of products of Q's such that different groups have no common suffixes whilst within each group the triads are connected to each other by having common points, we shall say such a product 'contains p groups'. Moreover, the number of times such a term occurs will be a polynomial in n whose order is equal to the number of distinct suffixes occurring in the product. If in a group a suffix only appears once, the inconsistency of the triad containing it is unaffected by the remainder of the group and the expectation of the product of Q's in that group will be zero. It follows that in all those terms which contribute something non-negative to μ_{2m+1}, none of the groups can contain a suffix which appears only once. Therefore, since all terms which contain more than m groups will have at least one group consisting of a single Q, the expectation of such terms will be zero. It follows that μ_{2m+1} is a polynomial in n, of degree $3m+1 = \left[\text{integral part of } \frac{3}{2}(2m+1)\right]$ at most, whose coefficients depend on m only. But $\mu_2(d)$ is of degree 3 in n and so

$$\frac{\mu_{2m+1}}{\mu_2^{\frac{1}{2}(m+1)}} = O\left(\frac{n^{3m+1}}{n^{3m+\frac{3}{2}}}\right) = O(n^{-\frac{1}{2}}).$$

Now consider μ_{2m}. This is a polynomial in n, and our aim is to find the order and coefficient of the term of largest order. In the first place we need only consider terms with m or less groups, for if a term has more than m groups, one at least will consist of a single Q and the expectation of the term will be zero. Moreover as before, in each term, the suffixes in each group must each occur at least twice in that group. The number of times each type of term occurs will be a polynomial in n of order equal to the total number of distinct suffixes in that term. As we shall show the leading term in $\mu_{2m}(d)$ to be of order $3m$, we can neglect terms whose frequency is less than this and therefore we can neglect all terms in which a suffix appears more than twice. Now consider a term with fewer than m groups and therefore containing a group of

order greater than two in the Q's. As no suffix can occur more than twice, no Q can occur more than once. Consider any Q_{ijk}, say, of this group. Then either the suffixes i, j, k are common to three other triads or one, i, say, is common to another triad and j, k common to a third. In either case evaluation of the expectation shows it to be zero. We can therefore restrict our attention to the case where there are m groups each containing two triads. Such groups can only be of the form Q_{123}^2, $Q_{123}Q_{124}$, $Q_{123}Q_{145}$ and the expectations of the two latter are zero whilst the expectation of Q_{123}^2 is

$$\tfrac{1}{4}(1-\tfrac{1}{4})^2+\tfrac{1}{4}(\tfrac{1}{4})^2=\tfrac{3}{32}.$$

The number of groups is m and the number of ways of choosing m such distinct pairs out of $(\Sigma Q_{ijk})^{2m}$ is $\dfrac{(2m)!}{2^m m!}$ so that the leading term in μ_{2m} is

$$\frac{(2m)!}{2^m m!}(\tfrac{3}{32})^m n^{2m},$$

whilst the leading term in μ_2 is $\frac{3}{32}n^2$ and so

$$\frac{\mu_{2m}}{\mu_2^m}\to\frac{(2m)!}{2^m m!}.$$

The distribution therefore tends to normality.

REFERENCES

KENDALL, M. G. & BABINGTON SMITH, B. (1940). *Biometrika*, **31**, 324.
KENDALL, M. G. (1943). *The Advanced theory of Statistics*, **1**. London: Charles Griffin and Co.

Notes on the calculation of autocorrelations of linear autoregressive schemes

BY M. H. QUENOUILLE

1. Bartlett (1946) has recently shown how, for a series of observations, we can test whether the observations can be adequately represented by a linear autoregressive scheme

$$u_{n+j}+a_1u_{n+j-1}+\ldots+a_ju_n=\epsilon_{n+j}, \tag{1}$$

where the a_i are known or fitted values, and ϵ_{n+j} is an error component independent of u_{n+j-1}. Bartlett's test is based on the formula

$$\operatorname{cov}(r_s, r_{s+t})\sim\frac{1}{n-s}\sum_{i=-\infty}^{\infty}\rho_i\rho_{i+t},$$

where r_s is the estimate of the true autocorrelation ρ_s between u_i and u_{i+s}.

The purpose of the note is to demonstrate how, using generating functions, ρ_i and $\sum_{i=-\infty}^{\infty}\rho_i\rho_{i+t}$ can be calculated with the minimum of computation.

2. The method of generating functions seems to have been used by Wold (1938), who applied them to finding the variances and covariances of linear forms of finite extent in variables such as ϵ_{n+j}. We shall, however, be concerned with linear forms of infinite extent.

It can easily be shown that the solution of (1) can be written

$$u_n=\epsilon_n+b_1\epsilon_{n-1}+b_2\epsilon_{n-2}+\ldots, \tag{2}$$

where

$$(1+a_1t+\ldots+a_jt^j)^{-1}=1+b_1t+b_2t^2+\ldots. \tag{3}$$

For example, if

$$u_{n+2}+au_{n+1}+bu_n=\epsilon_{n+2},$$

we have

$$(1+at+bt^2)^{-1}=(1-2x\cos\theta+x^2)^{-1}=1+\frac{\sin 2\theta}{\sin\theta}x+\frac{\sin 3\theta}{\sin\theta}x^2+\ldots,$$

where

$$\cos\theta=-\tfrac{1}{2}a/\sqrt{b},\ x=t\sqrt{b},$$

and hence

$$b_l=b^{\frac{1}{2}l}\frac{\sin l\theta}{\sin\theta}=\frac{2b^{\frac{1}{2}l}}{\sqrt{(4b-a^2)}}\sin l\theta.$$

3. Using this generating function we have

$$\sigma^2 \sum_{i=-\infty}^{\infty} \rho_i t^i = \operatorname*{Lt}_{c \to 1} \left[\{1 + a_1 t + \ldots + a_j t^j\} \left\{1 + a_1 \frac{c}{t} + \ldots + a_j \left(\frac{c}{t}\right)^j\right\}\right]^{-1}. \tag{4}$$

Now the expansion of (4) can be achieved by splitting into partial fractions and, in general, we can let $c \to 1$ before this operation is performed. Thus

$$\frac{t^j}{(1 + a_1 t + \ldots + a_j t^j)(t^j + a_1 t^{j-1} + \ldots + a_j)} = \frac{A_0 + A_1 t + \ldots + A_{j-1} t^{j-1}}{1 + a_1 t + \ldots + a_j t^j} + \frac{B_j + B_{j-1} t + \ldots + B_1 t^{j-1}}{t^j + a_1 t^{j-1} + \ldots + a_j}, \tag{5}$$

and using $\rho_i = -\rho_{-i}$, we can see that

$$A_i = -B_j/a_j \quad (i = 0)$$

$$= B_i - a_i B_j/a_j \quad (i = 1, \ldots j-1).$$

Thus the autocorrelations will be generated by

$$1 + \frac{t}{A_0} \frac{B_1 + B_2 t + \ldots + B_j t^{j-1}}{1 + a_1 t + \ldots + a_j t^j} + \frac{1}{A_0 t} \frac{B_1 + B_2 t^{-1} + \ldots + B_j t^{-j+1}}{1 + a_1 t^{-1} + \ldots + a_j t^{-j}}, \tag{6}$$

where the first term is expanded in powers of t and the second term is expanded in powers of t^{-1}.

4. The expression (6) can now be squared to give a generating function for $\sum_{i=-\infty}^{\infty} \rho_i \rho_{i+t}$. It will be necessary to split

$$\frac{t(B_1 + B_2 t + \ldots + B_j t^{j-1})(B_1 t^{j-1} + B_2 t^{j-2} + \ldots + B_j)}{(1 + a_1 t + \ldots + a_j t^j)(t^j + a_1 t^{j-1} + \ldots + a_j)}$$

into partial fractions, but the labour will be reduced since the matrix of the coefficients of the equations in B_i will be unaltered.

5. To illustrate the method, we can consider Kendall's series 1, which was used by Bartlett in his example.

The autoregressive scheme for this series is

$$u_{n+2} - 1{\cdot}1 u_{n+1} + 0{\cdot}5 u_n = \epsilon_{n+2},$$

so that $$\sigma^2 \sum_{i=-\infty}^{\infty} \rho_i t^i = \frac{t^2}{(1 - 1{\cdot}1t + 0{\cdot}5t^2)(t^2 - 1{\cdot}1t + 0{\cdot}5)} = \frac{2B_2 + (B_1 + 2{\cdot}2B_2)t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \frac{B_2 + B_1 t}{t^2 - 1{\cdot}1t + 0{\cdot}5},$$

where

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} -3{\cdot}7692 & 2{\cdot}1154 \\ -2{\cdot}1154 & -1{\cdot}4423 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2{\cdot}1154 \\ -1{\cdot}4423 \end{bmatrix}.$$

Thus $$\sigma^2 = 2{\cdot}8846,$$

and

$$\sum_{i=-\infty}^{\infty} \rho_i t^i = 1 + t \frac{0{\cdot}7333 - 0{\cdot}5t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \frac{1}{t} \frac{0{\cdot}7333 - 0{\cdot}5t^{-1}}{1 - 1{\cdot}1t^{-1} + 0{\cdot}5t^{-2}}, \tag{7}$$

so that

$$\rho_i - 1{\cdot}1\rho_{i-1} + 0{\cdot}5\rho_{i-2} = 0 \quad (i > 0). \tag{8}$$

If we now consider the square of the expression (7) we have a product term

$$\frac{2t(0{\cdot}7333 - 0{\cdot}5t)(0{\cdot}7333t - 0{\cdot}5)}{(1 - 1{\cdot}1t + 0{\cdot}5t^2)(t^2 - 1{\cdot}1t + 0{\cdot}5)} = \frac{-2B_2 + (B_1 + 2{\cdot}2B_2)t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \frac{B_2 + B_1 t}{t^2 - 1{\cdot}1t + 0{\cdot}5},$$

where

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} -3{\cdot}7692 & 2{\cdot}1154 \\ -2{\cdot}1154 & -1{\cdot}4423 \end{bmatrix} \begin{bmatrix} -0{\cdot}7333 \\ 1{\cdot}5754 \end{bmatrix}$$

$$= \begin{bmatrix} 0{\cdot}5686 \\ -0{\cdot}7210 \end{bmatrix},$$

and, if we write $$P_i = \sum_{i=-\infty}^{\infty} \rho_i \rho_{i+i} \Big/ \sum_{i=-\infty}^{\infty} \rho_i^2,$$

$$\sum_{i=-\infty}^{\infty} \rho_i^2 \sum_{i=-\infty}^{\infty} P_i t^i = 1 + 2t \frac{0{\cdot}7333 - 0{\cdot}5t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + t^2 \left(\frac{0{\cdot}7333 - 0{\cdot}5t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} \right)^2$$

$$+ 1{\cdot}4420 + t \frac{0{\cdot}5686 - 0{\cdot}7210t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \text{terms in } t^{-1}$$

$$= 2{\cdot}4420 + t \frac{2{\cdot}0352 - 1{\cdot}7210}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + t^2 \frac{0{\cdot}5377 - 0{\cdot}7333t + 0{\cdot}25t^2}{(1 - 1{\cdot}1t + 0{\cdot}5t^2)}$$

$$+ \text{terms in } t^{-1}. \tag{9}$$

From this we have $\sum_{i=-\infty}^{\infty} \rho_i^2 = 2{\cdot}4420$ and the 'correlations' P_i of the correlations are 0·8334, 0·4321, 0·0006, ... Successive terms may be calculated using the relation

$$P_i - 2{\cdot}21 P_{i-1} + 2{\cdot}2 P_{i-2} - 1{\cdot}1 P_{i-3} + 0{\cdot}25 P_{i-4} = 0 \quad (i > 0). \tag{10}$$

The calculation of $\sum_{i=-\infty}^{\infty} P_i^2$, suggested by Bartlett, can also be made by this method, but it is more arduous, and the first few terms will give a good approximation.

7. The same method can be used to calculate the appropriate number of degrees of freedom for testing the correlation between two linear autoregressive schemes.

In general, if $E(u_i u_j) = \rho_{ij}\sigma^2$, $E(v_i v_j) = \rho'_{ij}\sigma'^2$ and $r = \dfrac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\left(\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 \right)}}$,

then $$\operatorname{var} r \sim \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} \rho'_{ij}}{\sum_{i=1}^{n} \rho_{ii} \sum_{i=1}^{n} \rho'_{ii}}.$$

For linear autoregressive schemes, $\rho_{ij} = \rho_{i-j}$, $\rho'_{ij} = \rho'_{i-j}$, and thus

$$\operatorname{var} r \sim \frac{n + (n-1)\rho_1\rho'_1 + \ldots + \rho_{n-1}\rho'_{n-1}}{n^2}$$

$$\sim \sum_{i=-\infty}^{\infty} \rho_i \rho'_i / n.$$

Thus, provided n is large, r can be tested with $n \Big/ \sum_{i=-\infty}^{\infty} \rho_i \rho'_i$ degrees of freedom, and the calculation of $\sum_{i=-\infty}^{\infty} \rho_i \rho'_i$ can be made by the above method.

8. Finally, it is worth noting that, for autoregressive schemes involving m observables, it is possible to extend this method by the use of m parameters to calculate the correlations within and between the observables, provided that adequate estimates of the coefficients of the equations are available. In practice, however, the procedure will often be reversed, and estimates of the coefficients of the autoregressive schemes will be obtained by equating the theoretical and observed correlations.

REFERENCES

BARTLETT, M. S. (1946). *J.R. statist. Soc. Suppl.* **8**, 27.
WOLD, H. (1938). *A Study in the Analysis of Stationary Time-series*. Uppsala.

Approximate formulae for the percentage points of the incomplete beta function and of the χ^2 distribution

By D. HALTON THOMSON

Valuable 'Tables of percentage points of the Incomplete Beta Function' have been published in *Biometrika* (Thompson, 1941*a*) giving numerical values of percentage points at various probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$ for degrees of freedom $\nu_1 = 2q$ and $\nu_2 = 2p$ ranging up to 120, and with an accuracy of five significant figures. In the same volume, a 'Table of the percentage points of the χ^2 distribution' was also published (Thompson, 1941*b*) for values at the same probability levels and degrees of freedom ranging up to $\nu = 100$, and with an accuracy of six significant figures, thus supplementing the table of that function originally due to R. A. Fisher (Fisher & Yates, 1938).

Cases arise in practice where the tails of the frequency distribution of a large population are of special interest, thus involving (in the case of the beta-function) values of $2p$ larger than 120, with a small $2q$, or vice versa. Harmonic interpolation between 120 and infinity, however, leads to substantial errors, as is found when the values of the percentage points x are expressed in terms of their tail values (x or $1-x < 0{\cdot}5$). This Note shows that close approximations to such extreme values may be determined by using the χ^2 table as an auxiliary table to extend the Beta Function Tables in conjunction with certain simple alternative formulae. Comparisons within the range of the published Beta Function Tables are made indicating the degree of accuracy within that range. The accuracy of these formulae beyond that range increases rapidly with increasing $2p$ and decreasing $2q$ (and vice versa), so that they can be applied with confidence under such conditions.

The 'normalized' form of the Incomplete Beta Function, in the usual notation, is

$$P = I_x(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} \int_0^x x^{p-1}(1-x)^{q-1}\,dx, \tag{1}$$

in which, for a given P, $1-x(q,p)$ in the tables denotes the upper percentage point and $x(p,q)$ the lower percentage point.

It is known that, when p is large and q is small compared with p, this form tends towards the Incomplete Gamma Function

$$P = \frac{p^q}{\Gamma(q)} \int_0^t e^{-pt} t^{q-1}\,dt,$$

where $x(p,q) = e^{-t}$. This in turn may be transformed to the χ^2 distribution by putting $pt = [\chi^2_{2q}(P)]/2$. For a given large p and small q, therefore, the percentage point in terms of χ^2 is given approximately by

$$x(2p,2q) \cong \exp\left[-\frac{\chi^2_{2q}(P)}{2p}\right], \tag{2}$$

where $2q = \nu$ in the χ^2 table. This expression gives the exact value of x, when $2q = 2$, but for larger $2q$ the error, which is consistently negative, increases rapidly with increasing $2q$ unless $2p$ is very large—much larger than 120. It is, therefore, of limited practical use. The following modifications were in consequence evolved.

Approximation A

Consider the constant of integration in the original form (1) which, when expanded, is

$$\frac{(p+q-1)(p+q-2)\ldots(p+1)p}{\Gamma(q)}.$$

Let the terms $q-1, q-2, \ldots, 1, 0$ be averaged; the constant as a first approximation then becomes

$$\frac{\{p+\tfrac{1}{2}(q-1)\}^q}{\Gamma(q)}.$$

The numerator suggests that a more accurate approximation for x would be obtained by substituting $p+\frac{1}{2}(q-1)$ in place of p in (2), thus leading to

$$x(2p,2q) \cong \exp\left[-\frac{\chi^2_{2q}(P)}{2p+q-1}\right]. \tag{A}$$

A comparison of the approximate values of x obtained from (A) with the exact values in the Beta Function Tables, for all probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$, shows that:

(*a*) The error is consistently positive, but much smaller than the negative error in (2); in other words the latter is slightly over-corrected.

(*b*) For a given p/q and varying P, the error is nearly constant; it is smallest at $P = 0{\cdot}995$ and increases gradually in the direction of $P = 0{\cdot}005$.

(*c*) For a given P, the error decreases rapidly with increasing $2p$ and/or decreasing $2q$.

(*d*) Provided that p/q is larger than 4, the value of x is within 0·5 % of the exact tail value; if p/q is larger than 10, the error is within 0·1 % of that value.

Approximation B

The exponent in (A) may be written

$$\frac{\chi^2_{2q}(P)}{2p+q-1} = \frac{\chi^2_{2q}(P)}{2q}\,\frac{2q}{2p+q-1}$$

$$= \frac{\chi^2_{2q}(P)}{2q}\left[\frac{2}{2\left(\dfrac{2p+2q-1}{2q}\right)-1}\right].$$

The factor in square brackets is equivalent to the first term in the known expansion of the form

$$\log\left(\frac{n-1}{n}\right) = -2\left\{\frac{1}{2n-1}+\frac{1}{3(2n-1)^3}+\cdots\right\},$$

where $n = (2p+2q-1)/2q$, which converges rapidly when n is large; i.e. when $2p$ is large compared with $2q$. The above exponent may therefore be written

$$-\frac{\chi^2_{2q}(P)}{2q}\log\left(\frac{2p-1}{2p+2q-1}\right),$$

which, when inserted in (A), leads to $$x(2p, 2q) \cong \left(\frac{2p-1}{2p+2q-1}\right)^k, \qquad \text{(B)}$$

where $k = [\chi^2_{2q}(P)]/(2q)$.

A similar comparison with the Beta Function Tables, for the same range of probability levels, shows that:

(*a*) Approximation (B) gives generally more accurate values than (A), except when $2q$ is very small, in which case they are nearly identical.

(*b*) For a given p/q and varying P, the error is negligible in the vicinity of $P = 0{\cdot}25$; it increases negatively in the direction of $P = 0{\cdot}995$, and positively in the direction of $P = 0{\cdot}005$, the largest errors occurring at this level.

(*c*) For a given P, the error decreases rapidly with increasing $2p$ and/or decreasing $2q$.

(*d*) Provided that $(2p)^3/(2q)^2$ is larger than about 150, the values of x are within 0·5 % of the exact tail value; this implies that if $2p$ is larger than about 150, this degree of accuracy is attained even when p/q is as low as unity. If $(2p)^3/(2q)^2$ is larger than about 2000, the error is within 0·1 % of the exact value, which implies that if $2p$ is larger than about 120, this degree of accuracy is attained when p/q is as low as 4.

It will be observed that, when $2q = 2$, the formula does not revert exactly to (2), as is required by theory; but, unless $2p$ is also quite small, the error in the computed value of x is negligible.

The expansion of (B) leads to $$x(2p, 2q) \cong e^{-v} - s(1+\tfrac{1}{2}s)v + sv^2,$$

where $$v = \frac{\chi^2_{2q}(P)}{2p+2q-1} \quad \text{and} \quad s = \frac{q}{2p+2q-1},$$

thus demonstrating its analogies with Campbell's formula (C) below.

Adaptation of Campbell's formula

In a book concerned primarily with quality control, Simon (1941) quotes (without the proof) a formula, due to Campbell (1923), designed to determine the average number of defectives in a sample of n, starting from the known average number in an infinite sample. It is a particular application of the general problem now under consideration, namely, the approximate determination of the percentage points of

the Beta Function, starting from the corresponding known values for the χ^2 form of the Poisson exponential binomial summation. It is given in the following form:

$$\frac{a(c,n,P)-a(c,\infty,P)}{a(c,\infty,P)} = An^{-1}+\tfrac{1}{12}[14A^2+(3a+2)A+a]n^{-2}+\ldots, \qquad (3)$$

where $a(c,n,P)$ = average number of defectives in which P is the probability of at least c defectives in a sample of n, $a(c,\infty,P)$ = average number of defectives in an infinite sample, $A=\frac{1}{3}(c-a-1)$, in which $a=a(c,\infty,P)$. (Simon quotes $a=(a,\infty,P)$, which is an evident misprint.)

If G denotes the value given by the formula, then

$$a(c,n,P)=a(c,\infty,P)(1+G),$$

so that $1+G$ is the factor by which the average number of defectives in an infinite sample must be multiplied to give that in a sample of n.

The change from Campbell's notation to the more familiar general notation is given by

$$a(c,n,P)=\{1-x(2p,2q)\}n, \quad a=a(c,\infty,P)=[\chi^2_{2q}(P)]/2,$$

where $$n=p+q-1, \text{ and } c=q.$$

Let $$u=a/n \quad \text{and} \quad r=(c-1)/(2n),$$

then $$A=n(r-u/2).$$

By inserting this notation in (3) and rearranging the terms, the formula leads to

$$x(2p,2q)\cong 1-\left\{1+r\left(1+\tfrac{7}{6}r+\frac{1}{6n}\right)\right\}u+(1+\tfrac{11}{6}r)\frac{u^2}{2}-\tfrac{1}{6}u^3+\ldots,$$

which expression includes the first four terms in the expansion of e^{-u}.

Hence, for the determination of the percentage points $x(2p,2q)$, Campbell's formula may, in effect, be re-written as

$$x(2p,2q)\cong e^{-u}-r\left(1+\tfrac{7}{6}r+\frac{1}{6n}\right)u+\tfrac{11}{12}ru^2, \qquad (C)$$

where $$u=\frac{\chi^2_{2q}(P)}{2(p+q-1)} \quad \text{and} \quad r=\frac{q-1}{2(p+q-1)}.$$

For large $2p$ and small $2q$, the last two terms become negligible, in which case it reduces to

$$x(2p,2q)\cong e^{-u}-r(1+\tfrac{7}{6}r)u. \qquad (C')$$

Cochran's approximation

Cochran (1940), extending a method of Fisher's (1925), has introduced a useful approximation for the percentage points of the Incomplete Beta Function, *when both p and q are large*, his method being to determine a sufficiently accurate value of z, as used in Fisher's z-transformation.

If y is the normal deviate at probability level P, then for a given pair of arguments $2p$, $2q$, the following are first calculated, using Hartley's (1941) notation:

$$\lambda=\tfrac{1}{6}(y^2+3), \quad A=\frac{8pq}{2p+2q},$$

$$z=\frac{y}{\sqrt{(A-\lambda)}}+\frac{(\lambda-\frac{1}{6})(A-2p)}{pA}.$$

Hence, by Fisher's transformation, $$x(2p,2q)\cong\frac{2p}{2p+2qe^{2z}}. \qquad (D)$$

Comparison of formulae

Table 1 compares the various formulae for upper percentage points at an extreme probability level ($P=0{\cdot}995$). Table 2 indicates their relative accuracy on a common basis, namely, as a percentage of the exact value of x or $1-x$, whichever is the smaller, so that the deviations from the exact values, when x or $1-x$ approach zero, are duly emphasized. For intermediate probability levels, the percentages lie between the tabulated extremes. It will be noted that in the case of approximation B the errors pass through zero near the mid-range of P; in the cases of A and C the errors are positive for all values of P.

The general conclusions from these tables and other comparisons are that, for a given probability level P:

(*a*) When $p/q >$ about 6, approximations A, B and C have about the same degree of accuracy, so that the simpler, A or B, have the advantage.

(*b*) In the range $6 > p/q > 4$, there is little to choose between B and C; but B is the simpler.

(*c*) When $p/q < 4$ and the distribution approaches symmetry, D gives the best results, provided that $2p$ and $2q$ are moderately large, say > 50. It may be, however, that B in this range will be sufficiently accurate for many purposes; if $p/q > 2$, the maximum error of x is about 2 units in the third decimal place.

Table 1. *Comparison of approximate formulae at a given probability level*

$P = 0{\cdot}995$

$2p$	$2q$	$x(2p, 2q)$				
		A	B	C (Campbell)	D (Cochran)	Exact
120	2	$0{\cdot}9^4 16461$	$0{\cdot}9^4 16459$	$0{\cdot}9^4 16461$	0·999862	$0{\cdot}9^4 16461$
		Nil	*$-0{\cdot}0^4 00002$*	*Nil*	*−0·000054*	—
	4	0·9982908	0·9982906	0·9982907	0·997926	0·9982907
		+0·0000001	*−0·0000001*	*Nil*	*−0·000355*	—
	10	0·982764	0·982755	0·982760	0·982076	0·982759
		+0·000005	*−0·000004*	*+0·000001*	*−0·000683*	—
	20	0·944002	0·943893	0·943941	0·943366	0·943930
		+0·000072	*−0·000037*	*+0·000011*	*−0·000564*	—
	30	0·902230	0·901839	0·902000	0·901551	0·901950
		+0·000280	*−0·000111*	*+0·000050*	*−0·000399*	—
	40	0·86160	0·86070	0·86106	0·86066	0·86093
		+0·00067	*−0·00023*	*+0·00013*	*−0·00027*	—
	60	0·78782	0·78522	0·78621	0·78568	0·78579
		+0·00203	*−0·00057*	*+0·00042*	*−0·00011*	—
	120	0·62598	0·61430	0·61855	0·61620	0·61620
		+0·00978	*−0·00190*	*+0·00235*	*Nil*	—

N.B. The figures in italics are the differences between the approximate and exact values.

Table 2. *Relative accuracy of approximate formulae*

Error of $x(2p, 2q)$ expressed as a percentage of the smaller exact tail value (x or $1-x < 0{\cdot}5$)

$2p$	$2q$		A			B			C (Campbell)			D (Cochran)		
		p/q \ P	0·995	0·500	0·005	0·995	0·500	0·005	0·995	0·500	0·005	0·995	0·500	0·005
			%	%	%	%	%	%	%	%	%	%	%	%
120	12	10	*	*	*	*	*	*	*	*	*	−2·8	−0·1	+0·3
	20	6	+0·1	+0·2	+0·2	−0·1	*	+0·1	*	*	+0·1	−1·0	*	+0·1
	40	3	+0·5	+0·6	+0·7	−0·2	*	+0·2	+0·1	+0·1	+0·3	−0·2	*	*
	60	2	+0·9	+1·1	+1·2	−0·3	*	+0·2	+0·2	+0·1	+0·5	−0·1	*	*
	120	1	+2·5	+2·7	+4·4	−0·5	*	+0·7	+0·6	+0·5	+1·7	*	*	*
30	3	10	*	*	+0·1	*	*	*	*	*	+0·1	−30·0	−1·4	−7·1
	5	6	+0·1	+0·1	+0·3	−0·1	−0·1	+0·1	*	*	+0·4	−11·1	−0·5	−1·8
	10	3	+0·4	+0·5	+0·8	−0·3	−0·1	+0·3	+0·1	+0·1	+1·3	−2·2	−0·1	−0·5
	15	2	+0·8	+1·0	+1·9	−0·5	−0·1	+0·6	+0·3	+0·1	+2·8	−0·7	*	−0·3
	30	1	+2·4	+2·7	+7·1	−1·0	−0·1	+1·8	+1·0	+0·5	+8·2	*	*	*

* Error smaller than ±0·05 %.

Wilson-Hilferty approximation for χ^2-adjustment

This formula (Wilson & Hilferty, 1931) for the percentage points of the χ^2 distribution is

$$\chi^2(P) = \nu\left\{1 - \frac{2}{9\nu} + y_P\sqrt{\frac{2}{9\nu}}\right\}^3,$$

where ν represents the degrees of freedom, and y_P the standardized normal deviate corresponding to probability level P. A table has been published in *Biometrika* (Merrington, 1941), comparing the approximations derived from this formula with the exact values, at various probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$. It shows the remarkable accuracy of the formula, the maximum errors varying from about $\pm 0{\cdot}04$, when $\nu = 30$, to about $\pm 0{\cdot}024$, when $\nu = 100$.

When these errors were plotted against the exact values on logarithmic paper, it was observed that for a given probability level, they varied inversely with $\sqrt{\nu}$ very closely. It follows that this square root relation may be used to adjust the Wilson-Hilferty formula, bringing the values computed therefrom still nearer to the exact values.

If the difference (at $\nu = 30$) between the Wilson-Hilferty value and the exact value, when multiplied by $\sqrt{(100/30)}$, is treated as a coefficient C (which may be positive or negative), the required adjustment for any value of ν is given by

$$\text{Adjustment} = C/\sqrt{\nu}.$$

For various probability levels P, the values of C are given in the following table:

P	C	P	C
0·995	+0·233	0·250	+0·039
0·990	+0·157	0·100	+0·056
0·975	+0·067	0·050	+0·035
0·950	+0·011	0·025	−0·015
0·900	−0·029	0·010	−0·120
0·750	−0·046	0·005	−0·227
0·500	−0·013	—	—

A test against the Merrington Table shows that this adjustment leads to values of χ^2, between $\nu = 30$ and $\nu = 100$ at all probability levels with an accuracy of $\pm 0{\cdot}001$, i.e. to four or five significant figures. Since the Wilson-Hilferty approximation assumes a normal distribution about $1 - 2/(9\nu)$, which tends to unity as ν increases to infinity, and since the adjustment tends to zero under those conditions, it follows that the latter may also be safely applied for an indefinitely large ν.

It should be added that an adjustment on similar principles is not applicable to the Fisher approximation for χ^2.

REFERENCES

Campbell, G. A. (1923). *Bell Syst. Tech. J.* January.

Cochran, W. G. (1940). Note on an approximate formula for significance levels of z. *Ann. Math. Statist.* **11**, 93.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd. 1st edition.

Fisher, R. A. & Yates, F. (1938), *Statistical Tables for Biological Agricultural and Medical Research.* Edinburgh: Oliver and Boyd.

Hartley, H. O. (1941). Tables of percentage points of the Incomplete Beta Function. Methods of interpolation. *Biometrika*, **32**, 166.

Merrington, M. (1941). Numerical approximations to the percentage points of the χ^2 distribution. *Biometrika*, **32**, 200.

Simon, L. E. (1941). *An Engineer's Manual of Statistical Methods*, p. 185. New York: John Wiley and Sons, Inc.

Thompson, C. M. (1941*a*, *b*). Tables of percentage points of the Incomplete Beta Function, *Biometrika*, **32**, 151.

Thompson, C. M. (1941). Table of percentage points of the χ^2 distribution. *Biometrika*, **32**, 187.

Wilson, E. B. & Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci. Wash.* **17**, 684.

REVIEWS

A First Course in Mathematical Statistics. By C. E. WEATHERBURN. Cambridge University Press. Price 15*s*.

An outstanding feature of the present statistical time is the number of text-books which are being written, and each one from a slightly different point of view. It is this which makes statistical theory interesting to study, for there can be no rigid approach to a subject which is used and expounded by so many and diverse persons. Professor Weatherburn has taken a rather formal mathematical exposition of the subject, and mathematical students will find his book both interesting and profitable to read. Numerical examples are given for the reader to apply the appropriate mathematical technique. It is possible that these would have been of greater utility if they had contained the material in its crude state, and had not been streamlined so that the application of the technique is immediately obvious, but nevertheless many new examples are there.

I am not sure whether this book will be entirely useful to students of other subjects than mathematics. While the mathematical analysis is undoubtedly clear it is possible that many will not be able to follow it in detail, and the conclusions of the analysis are not emphasized strongly. We may contrast with this Fisher's *Statistical Methods for Research Workers*, where no analysis is given, but where the relevant formulae and their interpretation are stated unmistakeably and their applications to material in its crude state set out so that the student may calculate for himself.

Probability theory is the foundation stone on which the whole of statistical theory is built. It is disappointing therefore to find that it is given somewhat perfunctory treatment in one chapter and the part it plays in (say) statistical tests of significance is not brought out and emphasized. There is a tendency nowadays in applying statistical technique to regard the 5 % and 1 % levels of significance as sacrosanct and those coming fresh to the subject should learn that custom is the only reason for their choice.

In spite of the criticisms which I make, however, I would recommend this book to students who have obtained some idea of the aims and objectives of statistical theory, and who are desirous of learning the development of the mathematical technique as well as its application. Professor Weatherburn's mathematical analysis makes pleasant reading and may well throw new light on old methods for those who have learnt the rudiments of the theory.

F. N. DAVID

Advances in Genetics, Volume 1. New York, N.Y.: Academic Press. 1947.

This is the first number of a new periodical, probably an annual, summarizing recent work in various fields. Of the nine articles, ranging from 12 to 96 pages, with mean 42·6, S.D. 7·89, and a positively skew distribution, perhaps the most interesting to European geneticists will be that on the genetics of the ciliate Protozoa, *Paramecium* and *Euplotes*. Here Sonneborn describes work almost entirely done in America, with very surprising results. Thus *Paramecium aurelia* consists of at least seven endogamous varieties, each with two exogamous mating types, which might be called sexes were it not that in *P. bursaria* one of the varieties has no less than eight mating types.

Shrode and Lush's article of the genetics of cattle gives a very condensed account of the large amount of work which has been done on the inheritance of economically important characters such as milk yield and growth rate. For example cattle biometricians have used the important concept of 'heritability', meaning the fraction of the variance of a character due to additive genetic differences. Within a herd this rarely exceeds 30 %. More space is devoted to work on the genetics of colour and the like, which is of far less economic importance, and the review of progeny testing methods is disappointingly brief. However, the bibliographical references will be useful. Similarly, Atwood's article on forage crops, though most valuable as a guide to the literature, does not give a detailed account of any of the biometric work which has been done on grasses and clovers.

Only two of the papers give data which a biometrician could immediately utilize. These are Gordon's account of polymorphism in fish populations, and Spencer's of mutations in wild *Drosophila* species, which unfortunately does not include some valuable recent Italian and Russian work. Gordon's results call for the development of methods of estimating gene frequency similar to those used with human blood groups. Spencer is mainly concerned with results, but these are often given in sufficient detail to interest biometricians, though no attempt is made to summarize Wright's fundamental statistical theory.

The other articles will be less attractive to biometricians, though it is of interest to see how statistical methods are demanded by the mere fact that the genus *Crepis*, whose evolution is reviewed by Babcock, includes 196 species, most of which have been examined cytologically, and between which 130 of the 38,220 possible crosses have been made.

The volume will be indispensable to geneticists. Biometricians certainly cannot neglect it.

J. B. S. H.

Mathematical Methods of Statistics. By H. Cramér, Princeton University Press. 1946. $6.00.

This book was written by Prof. Cramér during the war and has been published first in Sweden and then by an offset process by the Princeton University Press in the U.S.A. It is a definitive exposition of the theory of mathematical statistics as it existed in 1940 (about) and it is worth while therefore to consider its contents in some detail. Prof. Cramér has divided his exposition into three parts; the first part is purely mathematical. The theory of sets and of such Lebesgue measure as is necessary for the understanding of the second part is developed first of all. Such a development will be useful for the student of mathematical statistics coming fresh to the theory of measure in that he receives guidance as to what are the elements essential for him to understand. Chapters 11 and 12 on matrices determinants and quadratic forms and miscellaneous complements do not fit into this general scheme but have obviously been included here as part of the mathematical equipment necessary for the student. Possibly Chapter 10 on Fourier Integrals would have fitted more naturally into Part II but this is a matter of taste.

Part II begins with a formal development of the theory of probability as given by the French and Russian schools of probability, and which Prof. Cramér has already given in his Cambridge tract 'Random Variables and Probability Distributions'. The treatment here seems simpler, however, than in his earlier tract and there is a more practical flavour to his exposition. This part while still purely mathematical begins to introduce distributions and ideas which are familiar to the statistician.

The title of the third part is 'Statistical Inference' and the main outline is that of small sample theory developed during the past twenty-five years. The illustrations are numerical as well as mathematical and an attempt is made to show the student the numerical applications of the processes through which his mathematical theory leads him. The treatment is not exhaustive but the student who has assimilated this part will have little difficulty in extending his knowledge by further reading.

As a textbook of mathematical statistics this book will remain unrivalled for many years to come. The mathematical exposition is clear, the development of ideas logical throughout, and the theorems are presented in a very general way. Any student of mathematics who wishes to get a picture of what statistical theory is about will be led inevitably to a study of this book. To those who wish to become statisticians it will be necessary to supplement the reading by a practical course in which the mathematical tools are tried out on numerical examples. This aspect of statistical work the book does not cover, but it is obvious that this would be the case from the title. It only remains to say to the student 'This is a good book, buy it'.

F. N. DAVID

CORRIGENDA

(*Biometrika*, **34**, 176–7)

In J. Wishart's paper on 'The cumulants of the z and of the logarithmic χ^2 and t distributions', the following correction should be made:

p. 176, 1st line of section 3: *read* 'log | t |' *for* 'log t', in two places.

p. 177, 1st line following equation (32): *read* 'log | x |' *for* 'log x'.

(*All Rights reserved*)

BIOMETRIKA. Vol. XXXIV, Parts III and IV

CONTENTS

	PAGE
On the distribution of the rank correlation coefficient τ when the variates are not independent. By WASSILY HÖFFDING	183–196
The significance of rank correlations where parental correlation exists. By H. E. DANIELS and M. G. KENDALL	197–208
Testing for normality. By R. C. GEARY	209–242
The stratified semi-stationary population. By S. VAJDA	243–254
A simple approach to confounding and fractional replication in factorial experiments. By O. KEMPTHORNE	255–272
A comparison of stratified with unrestricted random sampling from a finite population. By P. ARMITAGE	273–280
Some theorems on time series. I. By P. A. P. MORAN	281–291
Rank correlation between two variables, one of which is ranked, the other dichotomous. By J. W. WHITFIELD	292–296
The variance of τ when both rankings contain ties. By M. G. KENDALL	297–298
A χ^2 'smooth' test for goodness of fit. By F. N. DAVID	299–310
An exact test for the equality of variances. By R. L. PLACKETT	311–319
The estimation from individual records of the relationship between dose and quantal response. By D. J. FINNEY	320–334
A power function for tests of randomness in a sequence of alternatives. By F. N. DAVID	335–339
A numerical solution of the problem of moments. By H. O. HARTLEY and S. H. KHAMIS.	340–351
Approximation to percentage points of the z-distribution. By A. H. CARTER	352–358
MISCELLANEA	
Note on the cumulants of Fisher's z-distribution. By LEO A. AROIAN	359–360
A note on the mean deviation from the median. By K. R. NAIR	360–362
On the method of paired comparisons. By P. A. P. MORAN	363–365
Notes on the calculation of autocorrelations of linear autoregressive schemes. By M. H. QUENOUILLE	365–367
Approximate formulae for the percentage points of the incomplete beta function and of the χ^2 distribution. By D. HALTON THOMSON	368–372
REVIEWS	
A First Course in Mathematical Statistics	373
Advances in Genetics	373
Mathematical Methods of Statistics	374

A volume of *Biometrika* contains about 400 pages, with plates and tables, and it is hoped that in future this will be published annually in two half-yearly issues.

Papers for publication should either be sent to

PROFESSOR E. S. PEARSON, Department of Statistics, University College, London, W.C. 1,

or if more convenient may be submitted through a member of the Editorial Committee, viz.

PROFESSOR HARALD CRAMÉR, University of Stockholm, Sweden.

DR R. C. GEARY, Statistics Branch, Department of Industry and Commerce, Dublin.

PROFESSOR M. GREENWOOD, F.R.S., London School of Hygiene and Tropical Medicine, London, W.C. 1.

PROFESSOR J. B. S. HALDANE, F.R.S., University College, London, W.C. 1.

DR G. M. MORANT, R.A.F. Institute of Aviation Medicine, R.A.F. Station, Farnborough, Hants.

DR JOHN WISHART, School of Agriculture, Cambridge.

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each.

The subscription price, *payable in advance*, is Inland 45*s*. *net* per volume and Abroad 54*s*. *net* (including packing and postage). Owing to the scarcity of early volumes, the following rates must now be charged for complete sets. Vols. I—XXXIV, including XX^B: £126. 5*s*. in wrappers, not including postage. At present certain volumes are out of print, but steps are being taken to re-issue these as quickly as printing facilities permit. Recent volumes may still be obtained at the wrapper price; this is 64*s*. inland, including postage. Index to Vols. I to V, 2*s*. *net*. Index to Vols. I to XV, 5*s*. *net*. Cheques must be made payable to *Biometrika*, crossed "*a/c Biometrika Trust*" and sent to The Secretary, *Biometrika Office*, Department of Statistics, University College, London, W.C. 1, to whom all orders for series, single copies and offprints should be addressed. All foreign cheques must be drawn in sterling and on a Bank having a London Agency.

First printed in Great Britain at the University Press, Cambridge
Reprinted by offset-litho by Percy Lund Humphries & Co., Ltd.